DOCUMENT RESUME

ED 127 922 IR 003 884

AUTHOR Bourne, Charles P.; And Others
TITLE Analysis of Errors in the University of California Union Catalog Supplement.
INSTITUTION California Univ., Berkeley. Inst. of Library Research.
REPORT NO ILR-7602
PUB DATE Jun 76
NOTE 81p.; Figures may reproduce poorly due to print size

EDRS PRICE MF-$0.83 HC-$4.67 Plus Postage.
DESCRIPTORS *Cataloging; *Error Patterns; *Library Research; Research Methodology; Tables (Data); *Union Catalogs; University Libraries
IDENTIFIERS University of California

ABSTRACT

A study examined the error rate of the "University of California Union Catalog Supplement," a 47 volume, computer produced book catalog of the materials cataloged by the nine University of California campuses during the years 1963-67. The study attempted not only to determine the rate and nature of errors, but also to develop a methodology for studying large bibliographic files. A stratified sample of 94 pages was examined, types and rates of errors were identified, and errors were categorized according to degree of seriousness. This report describes the methodology and provides tabular summaries and explanations of the results. (EMH)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED127922

ILR-7602

Analysis of Errors In the University of California Union Catalog Supplement

Charles P. Bourne
Margaret Buss
Deborah Sommer
Judy Todd

Institute of Library Research
University of California
Berkeley, California 94720

June 1976

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

IR003884

CONTENTS

		Page
ABSTRACT		vi
ACKNOWLEDGEMENTS		vii
TABLES		ii
FIGURES		iv
I.	INTRODUCTION	1
	A. Background	1
	B. Reasons for this Study	5
	C. Relevant Literature	5
II.	SUMMARY	8
	A. Methodology	8
	B. Results	8
III.	METHODOLOGY	10
	A. The Sample	10
	B. Method of Determining Error	10
	C. Error Categorization	14
	1. Type	14
	2. Location	18
	3. Effect	20
	4. Cause	26
	5. Language	27
	6. Non-Monographic Type	27
	7. Comments	28
	D. Method of Conversion of Collected Data into Machine Form for Tabulation	29
	E. Suggested Improvements in Methodology	30
IV.	RESULTS	32
	A. Introductory Comments	32
	B. Display and Discussion of Findings for the Author/Title Catalog	33
	C. Display and Discussion of Findings for the Subject Catalog	54
V.	REFERENCES	72
APPENDICES		
	A. List of Sample Pages	
	B. Example of a Sample Page Marked before Coding	
	C. Operational Definition of "Edition"	

TABLES

Table		Page
1.	Total Number of Errors Found in the Sample	33
2.	Summary of Computed Error Rate	34
3.	Number and Percent of All Errors in Sample by Type and Cause, Author/Title Catalog	35
4.	Number and Percent of All Errors in Sample by Type and Cause, Author/Title Catalog, Santa Cruz Records Only	37
5.	Number and Percent of All Minor Errors in Sample by Cause and Effect, Author/Title Catalog	39
6.	Number and Percent of All Minor Errors in Sample by Cause and Effect, Author/Title Catalog, Santa Cruz Records Only	39
7.	Number and Percent of All Serious Errors in Sample by Cause and Effect, Author/Title Catalog	41
8.	Number and Percent of All Serious Errors in Sample by Cause and Effect, Author/Title Catalog, Santa Cruz Records Only	42
9.	Number and Percent of All Fatal Errors in Sample by Cause and Effect, Author/Title Catalog	44
10.	Number and Percent of All Fatal Errors in Sample by Cause and Effect, Author/Title Catalog, Santa Cruz Records Only	44
11.	Number and Percent of All Errors in Sample by Location of Error Appearance and Cause, Author/Title Catalog	46
12.	Number and Percent of All Errors in Sample by Location of Error Appearance and Cause, Author/Title Catalog, Santa Cruz Records Only	48
13.	Number and Percent of All Errors in Sample by Type and Location of Error Appearance, Author/Title Catalog	49
14.	Number and Percent of All Errors in Sample by Location of Error Source and Cause, Author/Title Catalog	50
15.	Number and Percent of All Errors in Sample by Location of Error Source and Cause, Author/Title Catalog, Santa Cruz Records Only	52
16.	Number and Percent of All Errors in Sample by Type and Cause, Subject Catalog	55
17.	Number and Percent of All Errors in Sample by Type and Cause, Subject Catalog, Santa Cruz Records Only	57
18.	Number and Percent of All Minor Errors in Sample by Cause and Effect, Subject Catalog	58

Table		Page
19.	Number and Percent of All Minor Errors in Sample by Cause and Effect, Subject Catalog, Santa Cruz Records Only	58
20.	Number and Percent of All Serious Errors in Sample by Cause and Effect, Subject Catalog	60
21.	Number and Percent of All Serious Errors in Sample by Cause and Effect, Subject Catalog, Santa Cruz Records Only	61
22.	Number and Percent of All Fatal Errors in Sample by Cause and Effect, Subject Catalog	63
23.	Number and Percent of All Fatal Errors in Sample by Cause and Effect, Subject Catalog, Santa Cruz Records Only	64
24.	Number and Percent of All Errors in Sample by Location of Error Appearance and Cause, Subject Catalog	65
25.	Number and Percent of All Errors in Sample by Location of Error Appearance and Cause, Subject Catalog, Santa Cruz Records Only	65
26.	Number and Percent of All Errors in Sample by Type and Location of Error Appearance, Subject Catalog	67
27.	Number and Percent of All Errors in Sample by Type and Location of Error Appearance, Subject Catalog, Santa Cruz Records Only	68
28.	Number and Percent of All Errors in Sample by Location of Error Source and Cause, Subject Catalog	69
29.	Number and Percent of All Errors in Sample by Location of Error Source and Cause, Subject Catalog, Santa Cruz Records Only	70

FIGURES

Figure		Page
1.	Final Version of Error Code Categories	12
2.	Error Coding Data Sheet	13
3.	Example of Error Type 10--Duplicate Data Not Suppressed	14
4.	Example of Error Type 20--Variant Cataloging Practice	15
5.	Example of Error Type 30--Orthographic Inaccuracy	15
6.	Example of Error Type 43--Misfile on Associated Title	16
7.	Example of Error Type 45--Misfile on Non-English Article	16
8.	Example of Error Type 50--Data Field Missing	16
9.	Example of Error Type 60--Data Field Added; Effect 40--Wasted Space	17
10.	Example of Error Type 70--Inappppropriate Entry	18
11.	Example of Error Type 80--Other	18
12.	Example of Error in Which the Source of Probable Error is not Apparent from Sample Entries	19
13.	Main Entry Form of Sample Added Entries Shown in Figure 12	19
14.	Example of Effect 21--Entire Entry Misfiled in Same Column	21
15.	Example of Effect 24--Entry Without Heading Misfiled in Same Column	21
16.	Example of Effect 30--Content Uncertain	22
17.	Example of Effect 51--Entire Entry Non-Consolidation, Same Column	23
18.	Example of Effect 54--Only Body of Entry Non-Consolidation, Same Column	23
19.	Example of Effect 61--Non-Consolidation of Heading Only, Resulting in Misfile of Entire Entry, Same Page	23
20.	Example of Effect 64--Non-Consolidation of Entry Heading Only, Resulting in Misfile of Body of Entry Only, Same Page	23
21.	Example of Subject Heading Non-Consolidation	24

Figure		Page
22.	Example of Effect 7A—Subject Heading Non-Consolidation Resulting in Entire Entry Misfile, Same Column	25
23.	Example of Effect 7Q--Subject Heading Non-Consolidation Resulting in Entry Heading Non-Consolidation, Same Page	25
24.	Example of Effect 7X--Subject Heading Non-Consolidation Resulting in Entire Entry Non-Consolidation, Same Column	25
25.	Example of Effect 80--Appearance Only	26
26.	Example of Effect 90--Other or Unknown	26

ABSTRACT

This study examines the error rate of the University of California Union Catalog Supplement, a computer produced book catalog of the materials cataloged by the nine University of California campuses during the years 1963-1967. The catalog, published in 1973, consists of 47 volumes of approximately 850 pages each and is divided into a 31 volume Author/Title section and a 16 volume Subject section. This study attempts not only to determine the rate and nature of error in this particular catalog, but also to provide a general methodology for studying error rates in large bibliographic files, whether computer produced or not.

A stratified random sample of 94 pages (5,900 entries) was taken from the catalog. The pages were thoroughly examined by two of the authors and each error discovered was analyzed according to six aspects: type, location, effect, cause, language, and non-monographic type. A total of 4,338 errors were found in the sample. This represents an error rate of 46.1 errors per page or 0.74 errors per entry. The sample from the Author/Title section of the catalog (where main entries consist of a complete bibliographic record) contained 3,167 errors, or 0.88 errors per entry. The sample from the Subject catalog (where entries consist only of subject heading, author, short title, date, location code & call number) contained 1,171 errors, or 0.51 errors per entry. Errors were categorized according to the degree of seriousness of their effects: minor, serious, and fatal. Minor errors made up approximately one half of all those found. Serious errors made up about 43% and fatal errors totaled approximately 7% of the error found in the sample.

ACKNOWLEDGEMENTS

This study had its beginning in a systems analysis class in the University of California School of Librarianship at Berkeley. This classroom experience laid the groundwork for our understanding of the problem of the U.C. Union Catalog Supplement (UCUCS) error rate and for the design of our methodology. The Institute of Library Research provided the source materials and continued support during this study. We also thank our fellow students for sharing their ideas which surely contributed to the project. Particular thanks is due Douglas Ferguson who produced a design for a study such as the present one. We also wish to thank the following people for answering our many questions about how UCUCS was produced and for suggesting ways to improve our design: Arjun Ayer, Mike Berger, Jay Cunningham, Tom Hargrove, Bill Harrelson, Allan Humphrey, Larry Legard, Linda Nakamura, and Maxine Sansing. We are very grateful to Keith Blair for writing the program used in data reduction, to the U.C. Computer Center keypunchers for converting the data, and to the typists who made this draft into a legible report so that it has become information available for use by others. We also greatly appreciate the work of Borut Dekleva, Mary Hackney, Victoria Hanawalt, Ron Heckart, Patiala Khoury, and Lena Menashian in examining foreign language entries in the sample.

I. INTRODUCTION

A. BACKGROUND

The University of California Union Catalog Supplement (hereafter referred to as UCUCS) is a computer-produced book catalog of the monographic materials cataloged by the nine University of California campuses during the period 1963-1967.[1] The catalog was intended to serve not only as a finding tool but also as a complete bibliographic record of the items cataloged during that five-year period. It is divided into a 31-volume Author/Title catalog and 16-volume Subject catalog. All subject entries appear in the Subject catalog, and all other added entries and main entries appear in the Author/Title catalog. Added entries in the Author/Title section are in brief form and refer to the main entries which are in full bibliographic detail.

The first step involved in producing the catalog was the collection of the main entry catalog cards for monographs cataloged during 1963-1967 from eight of the nine U.C. campuses. (One of the campuses, Santa Cruz, sent its records in machine readable form on magnetic tape instead of in catalog card form.) The collection of cards was performed by the Institute of Library Research (ILR) over a period of years. Over 1.1 million catalog records were collected and processed, constituting approximately 750,000 unique titles. The collected records were inspected before processing in order to remove those containing non-Roman alphabet characters. The remaining records were given a unique serial number (including a campus code) and then microfilmed to guard against fire or other loss. One of the functions of the unique identification number was its subsequent use in verification of the accuracy of the keyboarding process. The numbered records were then sent to a commercial vendor for Optical Character Recognition (OCR) typing. They were not formatted or edited before they were sent to the keying vendor; the keyboarding was done from the original catalog records, unaltered except for the identification number which had been stamped in the upper right corner of each card. Some rather gross delineation of parts of the catalog record was done by the typists in the process of keying the records. Their instructions were simply to key all the characters consecutively in order of appearance on the card, starting with the identification number, continuing with the call number and then the text of the record. The major data regions of the record were to be delineated by the typists by means of slashes ("/"), cross hash marks ("#") and plus signs ("+"). A slash was to be keyed at the start of each new paragraph indention on the card (data elements so marked included entry heading, start of title statement, collation, notes, tracing, etc.); a cross hash mark was to be keyed after the call number; and a plus sign was to be added at the end of the record. The typists were not expected to possess prior knowledge of the structure of a bibliographic record (other than whatever familiarity they may have had resulting from their own personal use of libraries), and were instructed to recognize the data elements purely on the basis of their locations on the card. (A copy of the instructions for keying the records may be obtained from ILR.)[2]

Since the UCUCS catalog was to include records of interest to many libraries in the academic community, it was important that the records be in MARC II communications format. An Automatic Format Recognition (AFR) program was developed by ILR to tag the bibliographic data fields of the records according to a subset MARC II format after they were keyed and returned to ILR in pre-AFR magnetic tape form. (For a description of the AFR program, see "Automatic Format Recognition," by Liz Gibson.[3] For a comparison of U.C.'s AFR program with other methods, see Brett Butler's article, "Automatic Format Recognition of MARC Bibliographic Elements: A Review and Projection."[4]) These programs were being written at the same time that the Format Recognition programs were being developed by LC. There were no programs in operation at that time that could have been used by U.C. instead of developing their own programs. The AFR program depended on the accuracy of the insertion of the slash and cross hash delimiters by the typists, as well as the accuracy of the typing itself. The AFR program relied not only on the data element delimiters but also on the content of the data fields. The keyboarding stage of the process was, therefore, very important for the success of the AFR program, and it was upon this program that subsequent computer manipulations of the records depended (sorting, merging, consolidation, etc.).

Although keyboarding accuracy was recognized as being very important, there was no key verification stage in the processing of UCUCS records. However, the keying vendor's contract stipulated a keying error rate not to exceed .5% (.005) of the keystrokes. According to the contract with the vendor, if this error rate were to be exceeded for any batch of records, either the vendor would rekey the batches which showed an excessive amount of error or else receive a reduced payment for the erroneous batch. The agreed upon method of determining the number of errors in a record and the error rate in batches of records was as follows. It was estimated that each record contains an average of 400 characters, excluding the delimiter characters to be added by the keyboarders. In determining the number of errors in a record, the text was divided into five-character strings. Each group of five characters was examined as a unit. If no error occurred in a five-character unit, then that unit was considered error free; if <u>any</u> error occurred in a five-character unit--regardless of what sort of error occurred. or how many actual keystrokes were involved--then that unit was considered to contain one error. Each record contained an average of about 80 five-character units. Five such records would be made up of about 400 units. The .5% error rate tolerance therefore meant that ILR would tolerate an average of two errors in 400 units or five records. In determining the vendor's error rate, ILR operations staff members took 3% random samples of records in each batch of 1,000 keyed records. These were printed out from the magnetic tape and compared with images of the original source records stored on microfilm. Errors were counted as described above. Batches which were determined by this sampling inspection to have a keying error rate in excess of .5% were to be returned to the vendor to be rekeyed.

However, the rekeying compounded the quality control problem still further. Subsequent sampling showed that the error rates of many of the rekeyed batches were even higher than the error rates prior to rekeying. Moreover, shipping the boxes of cards back and forth to the vendor (located in Ohio) took a great deal of time.

Due to time delays and the fact that the error rates of the recycled batches of records were as bad as or worse than those of the original keying attempts, the practice of returning to the vendor batches containing excessive error was soon discontinued. Instead, ILR accepted whatever was produced by the vendor and reduced payment proportional to the error rate in accordance with the stipulations of the contract. This choice was dictated almost entirely by time constraints on UCUCS production. If the contracted date when the data base was to be delivered to the photo-composition vendor could have been extended, it might have been possible to absorb the time delays created by shipping defective batches of records back and forth to the keying vendor. Since this was not possible, the actual keying error rate for the full UCUCS data base was bound to be in excess of .5%. The actual keying error rate was finally estimated to be an average of 0.54%, according to the operational definition of "error rate" agreed upon by ILR and the keying vendor.

The effort to control the quality of the keyboarding was one of several planned methods of quality control which were either only partially successful or else never implemented at all because of various influencing factors. It was never intended by the Project Manager that the data base be exhaustively manually proofread or edited. To do so would have been impossible given the then-prescribed time and monetary constraints of the UCUCS project. Moreover, a major purpose of the project, according to the Project Director, was to experiment with producing a book catalog with a minimum of manual intervention and with an error level that was supposedly agreed in advance as one of the product specifications. As much of the catalog production as possible was to be performed by the computer processes with only limited human inspection, including quality control. Three other major quality control efforts were originally planned, all to be executed mechanically with a minimum of human intervention. Only one of these was actually utilized, however, and it was used on only part of the data base.

It was intended that the Harvard Shelflist, available in machine-readable form, be used to derive a baseline name authority list of about 260,000 entries against which all author names in the UCUCS data base would be matched. Exact matches would be assumed to be spelled correctly. Author names which did not find a match in the authority list would be printed out for a manual check. This plan was never implemented because the software was incomplete by the time photocomposition began.

It was also intended that all English language words in title fields would be checked against an authority list. The authority list in this case was the shorter Oxford English Dictionary, available in magnetic tape form, with a total of about 75,000 words. A program to perform this matching operation was written by a Research Assistant who was a doctoral student in the Electrical Engineering and Computer Sciences Department at Berkeley. This program operated on the assumption that exact matches with the authority list were spelled correctly and that non-matches were misspelled. Words thus identified as misspelled would be printed out. In addition, the computer would attempt to predict the correct spelling and this too would be printed out. If visual inspection of the printout showed that no change was necessary, the proofreader would make no textual changes and merely submit a control card. After the next input of the data the computer would automatically accept the word it had initially identified as being misspelled. If the proofreader determined a different spelling of

the word to be correct, this spelling would be noted by the proofreader and then keypunched and fed into the computer to override the "automatic" correction. In effect, the computer program attempted to automatically correct the entries, subject to manual override and control. The program to perform this operation was written, and much of the debugging had been done. Unfortunately, the program was never used, due primarily to time constraints. Also, because the program was so complex and sophisticated, it was not certain that there was sufficient money available to pay for the computer time which would be necessary if it had been used in production.

A third mechanical quality control effort was used in production to a limited extent. This involved comparing the subject headings which would appear in the Subject catalog with a machine-readable authority list. The authority list in this case consisted of the machine file equivalents of LC Subject Headings List, 7th edition and the first annual supplement, plus approximately 50,000 subject headings, verified as legitimate U.C. headings not included in the U.C. list, plus about 3,000 of the geographic headings used in the U.C. Berkeley Library main catalog.[5] This comprised an authority list of about 120,000 entries. Similar to the title proofing program discussed above, this program would print out non-matches along with the near-matches which the program determined to be the correct form of the misspelled subject heading. The computer's suggested change would be made automatically if the proofreader did not override it and replace it with another spelling of that subject heading. In this process a number of variant forms of subject headings (such as "...Descr. and Trav." instead of "... Description & Travel") were normalized so that more entries would be brought together under one authenticated form of the subject heading. All of the subject headings in the total UCUCS file were run through the machine process. Approximately 40% of the needed changes, amounting to 1,058,072 changes, were ultimately made in the UCUCS Master File--again, limited because of time and monetary constraints.

All other operations performed to produce UCUCS were done by computer rather than manually. After the records were keyed and automatically tagged by the AFR program, duplicate entries were automatically consolidated into a machine-readable card set equivalent created for each union record, including all appropriate subject and added entries. The file was then exploded into all those entries which would comprise the Author/Title catalog and all those which would make up the Subject catalog. Finally the entries in each part of the catalog were sorted and merged into sequence, and the two parts of the catalog went through the line-and-column makeup, then page composition and printing process.

In summary, most of the source records went all the way from keyboarding to page printing and binding without manual editorial intervention at any point in the production cycle. This was a management decision that resulted in a high error rate (mostly because it was not possible within time and budgetary constraints to implement several of the planned programs and procedures) but with a lower unit cost than had been experienced by any other equivalent book catalog conversion effort before or since UCUCS. The unit cost of approximately $1.16 per record includes such processes as keyboarding and optical-scan reading of source records; formatting of source records; development and operation of an authority control file improvement system; consolidation of duplicate entries; formatting for videocomp composition;

and composition, printing, and binding of 250 copies of a catalog of 42,000 pages.

B. REASONS FOR THIS ERROR STUDY

There are five reasons for the present study.

i) UCUCS has been the object of much criticism due to the many errors in it. Much of the criticism has come from U.C. librarians who would like to rely on such a tool in their daily work. This criticism demonstrated the need for an objective measurement of the errors in the catalog. A goal of the present study was to provide a definitive statement of the nature and extent of the error in UCUCS. Consequently, both the rate and types of errors have been carefully analyzed in this study.

ii) A second reason for studying the errors in UCUCS was to improve the present machine-readable file from which the book form UCUCS was made. The file needs to be cleaned up for use by the U.C. Bibcenter for on-demand catalog card production and other applications that will use the UCUCS machine records. Moreover, the UCUCS file can serve as a valuable resource data base for subsequent use by libraries outside the University of California system. The fewer the errors in the file, the greater its value.

iii) The programs used to produce UCUCS also need improvement for use by the University-wide Library Automation Program for subsequent bibliographic processing efforts. A careful, definitive study of the errors in UCUCS can facilitate such program improvement.

iv) A fourth purpose of the study was to identify needed changes in local U.C. library procedures. Some of the error in the catalog is due to failures in preparation of the records prior to handling by the ILR staff. Some of the error is also due to variant cataloging practice among the nine U.C. campuses. This study is intended to help delineate areas where consideration could be given to improvement in local procedures.

v) A fifth important motivation for a study of this sort was a need for a general methodology for measuring errors in bibliographic catalogs--machine produced or otherwise--and in machine-readable bibliographic files. As far as we have been able to determine, no such generalized methodology has yet been designed. The present study analyzes both rates and types of errors in a particular catalog; in the process of doing this, a taxonomy of error types has been developed which should be applicable and useful in the error analysis of any bibliographic catalog or similar file.

C. RELEVANT LITERATURE

In searching library literature, the authors sought information regarding studies of errors in computer-produced book catalogs or other large catalogs. Although there are a few studies which consider the problem of error in such catalogs, none of them approaches the complexity of analysis attempted here.

The following literature was searched: Library Literature, 1967-June, 1974; Annual Review of Information Science and Technology, 1966-1973; and ERIC 1966-December 1974. The literature examined falls principally into three groups:

- Design-oriented or descriptive discussion
- Discussion of filing rules for machines and the problems caused thereby
- Evaluation of error in existing machine-readable printed bibliographic catalogs.

Much has been written regarding the design and implementation of machine conversion of files to book catalog format. Most of the literature seen either was of this type or described projects in process. The latter descriptions almost never included evaluations of error in the project discussed.

Some of the discussion about filing problems caused by computer filing is of value, since it helps explain the type and effect of a large number of the errors analyzed in this study. Cartwright's paper, "Mechanization and Library Filing Rules," discusses some of the filing problems which can occur in computer-produced book catalogs.[6] He suggests that filing problems should be considered at the time of file conversion, not after-the-fact. He also observes (correctly, in the authors' opinion) that standardized spacing and punctuation are essential, and that the lack of such standardization can lead to serious problems in filing and in the consolidation of entries.

Cartwright describes two book catalogs produced by computer: the book catalog at Florida Atlantic University and the Meyer Undergraduate Catalog at Stanford University. These systems had difficulty in filing entries in their proper order; for example, honorary titles such as "Sir" were used as filing elements, in contrast with the A.L.A. Filing Rules, which ignore titles in filing. No measurement of error is provided in this report.

A paper by Joe E. Hewitt is one of the few which actually reports and evaluates an error rate.[7] Hewitt points out the importance of performing catalog error studies not only for many uses by the local library staff but also for publication and use by other libraries in evaluating their own catalogs. He, like the present authors, decries the scarcity of reports of catalog error studies in library literature: "... the unfortunate result is that the research library performing an audit does so in a vacuum." His paper breaks this vacuum's seal by publishing a filing error rate of 1.1 percent in the University of Colorado's Norlin Library author/title card catalog of approximately 1,000,000 cards. Hewitt succinctly describes the general methodology used in arriving at this error rate and reports other statistical data. He also cites a 1953 audit of the official catalog of the Library of Congress reporting a filing error rate "in excess of 5 percent."

The third group of literature includes evaluations of errors in computer-produced catalogs. Some of these studies are solely concerned with cost-benefit analyses. Other studies evaluate the impact of a book catalog on the user, but error in such products is being ignored, or else no error exists.

Books in Print, 1969, a computer-produced catalog, has been analyzed for error in an article by N. Cambier, et al., "Books in Print 1969: An Analysis of Errors."[8] A sample of 2,000 entries was chosen; the BIP entries were then compared with Publishers' Trade List Annual to determine any discrepancies. After excluding factors such as foreign names and variant spellings, errors were typed as follows: Author omission, author error, title omission, title error, date omission, date error, price omission, defined; however, an error rate of 8.8 percent was reported. Although this study has a relatively simple methodology, it is nevertheless the study most comparable to the UCUCS error study.

Another study of some relevance to UCUCS is the one made by J.L. Dolby, et al., "Efficient Automatic Error Detection in Processing Bibliographic Records."[9] This report discusses an "error study" undertaken to determine the priority to be given procedures involving automatic error detection programs. Samples of unstated size and composition were drawn from Harvard University's Widener Shelf List and from Stanford University's Meyer Undergraduate Catalog. The samples from both libraries were proofread and corrections indicated on a computer output sheet. These sheets provided the data for the study.

No statement was made about the actual rate of error found. The analysis was aimed primarily at comparing the types of errors prevalent in the two samples and considering the possibility of automatic error detection. Errors were categorized into three groups:

- Sequence errors
- Missing information
  This group has six subdivisions, including: spaces, punctuation, capitalization, and diacritics.
- Incorrect information
  This group has nine subdivisions, including: proper names, words, punctuation, format, capitalization, and diacritics.

As can be seen, these categories are considerably broader than those used in the UCUCS study. Correlations were not made between type of error and cause of error, nor was an attempt made to determine the location of error within the entry.

II. SUMMARY

A. METHODOLOGY

A stratified random sample of 94 pages from UCUCS was used in this study --61 pages from the Author/Title section of the catalog and 33 from the Subject section. Each of these pages was Xeroxed and read by one of the authors who marked the copied page with red pencil and recorded each error found on an error coding sheet. Each page was re-examined by another one of the authors in order to catch any errors missed by the first person and also to insure continuous standardization of interpretation of the error codes. For each sample page, the two adjacent pages were also Xeroxed. These adjacent pages were not included in the sample in any way but were used to help find errors actually occuring on the sample page which would be impossible to notice without seeing the adjacent pages.

The analysts used an inclusive definition of error. An error was considered to be not only problems which caused loss of entry point or misfiling of entries, but also anything which might cause confusion or irritation on the part of the catalog user. Therefore, relatively minor mistakes such as improper spacing or print size were included as errors in this study. The "catalog user" was considered to include not only professional librarians but also students and the general public.

Although many kinds of errors were identified in the study, they were coded in such a way that they could be grouped later into three general categories of "fatal," "serious," and "minor;" thus relatively insignificant errors could be evaluated separately from more serious ones. Fatal errors were defined as those which would make it very likely that an entry point (i.e., a bibliographic record) would be lost to the user. Serious errors included non-fatal errors which would make it fairly likely that an entry point might be missed by the user and errors which render the content of the record unclear or misleading. Minor errors including those which merely affect the appearance of the entry without being likely to cause confusion for the user.

Each error found was coded according to six different aspects: Type, Location, Effect, Cause, Language, and Non-Monographic Type. Using each of these aspects in recording the errors made it possible to obtain a rather realistic and specific idea of the nature of the errors in UCUCS, and this, in turn, will enhance the efforts of programmers and systems analysts to improve the data base and the programs which produced the catalog.

After all of the errors had been coded, the collected data was keypunched, and most of the data reduction was done by computer.

B. RESULTS

A total of 4,338 errors were found on the 94 sample pages of 5,900 entries (3,589 entries in the Author/Title section and 2,311 in the Subject section). This represents an average of 46.1 errors per page, or 0.74 errors per entry. There were 3,167 errors in the Author/Title section and 1,171 errors in the Subject section. We can estimate that there are an

average of 0.88 errors per entry in the Author/Title section and 0.51 errors per entry in the Subject section. Generally, however, errors tend to "clump" in some entries rather than being evenly spread throughout the entries. Thus, many entries showed no errors. We can also estimate that there are approximately 51.9 errors per page in the Author/Title catalog and 35.5 errors per page in the Subject section.

If these figures seem a bit alarming, it must be remembered that they include all sorts of errors, even very minor ones. The serious and fatal errors together represent only about half of all the errors. The serious errors totaled 1,886 representing 43.5% of all found, and the fatal errors totaled 300, representing 6.9%. There were 2,152 minor errors in the sample, or about 50% of all the errors found.

III. METHODOLOGY

A. THE SAMPLE

Using the Rand Corporation Tables of Random Numbers, a stratified sample of 61 pages from the Author/Title catalog and 33 pages from the Subject catalog was selected.[10] The Author/Title catalog consists of about 26,900 pages in 31 volumes, and the Subject catalog consists of about 13,900 pages in 16 volumes. The sample selected represents an average of about two pages for each volume in the catalog, or approximately .23% of the entire sampling universe.

Each of these 94 pages (a list of the sample pages is given in Appendix A) was Xeroxed, along with the page immediately preceding and the page immediately following it. These two adjacent pages were not used as part of the sample but merely as aids in determining the errors actually occurring on the sample pages. In some cases an error would not be noticeable on an isolated page without referring to one of the adjacent pages. But in no case were errors appearing on the pages adjacent to the sample pages tallied in our study; only errors actually occurring on one of the 94 sample pages were analyzed and counted.

B. METHOD OF DETERMINING AN ERROR

An error may be an error may be an error, but what constitutes an error is still a matter of opinion. We have no doubt that what we have decided to include as an error may be considered by others as too trivial to be included or even not an error at all, and that we may have excluded instances of what others would term an error. This certainty arises from the experience of disagreement between the authors on the question of what should count as error, not to mention the disagreement within our respective minds at different times. The guiding principle which evolved through our discussion was to be as inclusive as possible within the realms of reason: to tally as an error everything from that which might cause even mild confusion or irritation on the part of the catalog user to those errors which are very serious and almost certainly result in a lost entry point. (Of course, we did not hold the Catalog responsible for those factors inherent in any bibliographic catalog which might be confusing to the user, such as arbitrary but typical nuances in filing rules.) "The user" was considered to include not only professional librarians but also students and the general public who might use the catalog. We included as error relatively minor typographical errors such as misplaced umlauts or accent marks, relatively minor spacing problems which might result in confusion or just difficult reading, such as the absence of a space in appropriate places in the collation statement (e.g., 261p. instead of 261 p.), and apparent inconsistencies in printing format which make the text difficult to read. Anticipating the disagreement with this inclusive policy, we have attempted to record our data and report our results as specifically as possible so that readers can more readily evaluate our conclusions in light of their own opinions and definitions of "error."

It should be noted that this inclusive policy of error determination will tend to make the overall error rate higher than would a less critical policy, and that it will also result in a relatively lower percentage of "fatal" and "serious" errors compared with the third category, "minor" errors. The authors defined fatal errors to be those which would make it very likely

that an entry point (i.e., bibliographic record) would be lost to the user. This category includes those errors which obviously involve a lost entry point and also entries which had been misfiled onto a page other than the one where they should have appeared (according to ALA filing rules). Serious errors are defined to include non-fatal errors which would either make it fairly likely that an entry point might be missed by the user (such as an entry misfiled on the correct page but in the wrong column) or errors which render the content of the record unclear or misleading. Minor errors include those which merely affect the appearance of the entry without being likely to cause confusion for the user.

To arrive at a consensus on what we wanted to include as errors and subsequently to ensure consistency in our use of the error categories, a duplicate set of 31 of the 94 sample pages (every third page) along with its two adjacent pages was Xeroxed. Two of the authors (Todd and Sommer) then examined a few of the pages independently. Each entry was numbered by starting with the first entry in each column as number 1 and proceeding through the remaining entries in that column. The errors were first marked in red pencil on the sample pages. (See Appendix B for illustration.) They were then coded on a data sheet according to our respective understandings of the first version of the error categories. The results were compared and discussed; revisions were made in the coding sheet, and definitions were clarified. More pages were similarily examined and results compared; more discussion and revision ensued. This process continued until an error categorization very near the present form was worked out. The remainder of the sample was then examined by both of these reviewers and errors were recorded. One further revision in the error categorization was subsequently made (an elaboration in the Effect categories 50 through 7Z described later). All entries coded according to earlier version were, of course, recorded according to the present form of the error categorization.

Each sample page (and its adjacent pages) was examined at least twice, once by each analyst. Of course the pages used in deriving the error categorizations were read in part more than twice. It was noticed that in looking over these pages more than twice, rarely were errors found that had been "missed" in previous examinations. It is therefore the opinion of the authors that examining each sample page once by each of two people is a reasonably satisfactory method to catch the vast majority of errors. Errors caught in subsequent readings will probably be minor and the effort of additional proofreading will not be worth the added expense.

Because the authors were not familiar with all the languages used in the catalog, a number of entries were examined by other people with the guidance of one of the authors. One of the present authors has an adequate reading knowledge of Spanish and Portuguese. Entries in many other langugages were checked by other ILR staff members and some Library School students. A total of 152 entries of our sample remained partially unanalyzed because they contained words in languages for which we had little linguistic capability (e.g., Turkish, Esperanto, Indonesian). However, even these entries were examined for errors which could be noticed without knowing the foreign language used. That is, most such entries contain some English words (at least in the tracings, notes, etc.), and it was not usually necessary to know the language in order to check for non-consolidation problems, extra blanks, etc.

I. Type
- 10 Duplicate data not suppressed
- 20 Variant form of same data or ed.
- 30 Orthographic inaccuracy
  - 31 transposition
  - 32 missing string
  - 33 added string
  - 34 meaningless string
  - 35 missing or added blank(s)
  - 36 incorrect or missing caps
  - 37 other misspelling or undetermined
- 40 String improperly used in filing
  - 41 function term
  - 42 dates
  - 43 associated title
  - 44 English article
  - 45 non-English article
  - 46 other
- 50 Data field missing
- 60 Data field added
- 70 Inappropriate entry
- 80 Other

II. Location
- 10 Main entry
  - 11 entry heading
  - 12 title statement
    - 1A (short) title
    - 1B subtitle
    - 1C author statement
    - 1D other
  - 13 edition statement
  - 14 place
  - 15 publisher
  - 16 date
  - 17 collation
  - 18 notes
  - 19 tracings
  - 1X call no./location
  - 1Y other
- 20 Added entry
  - 21 heading
  - 22 title statement
  - 23 date
  - 24 see reference
  - 25 call no./location code
  - 26 other
- 30 Subject entry
  - 31 subject heading
  - 32 entry heading
  - 33 title statement
  - 34 date
  - 35 call no./location code
  - 36 other

III. Effect
- 10 Lost entry point
- 20 Misfiled (not non-consol.)
  - 21 same col.)w/ 24 same col.)w/o
  - 22 same page)hdg. 25 same page)hdg.
  - 23 other) 26 other)
- 30 Content uncertain
- 40 Wasted space (not non-consol.)
- 50 Non-consol. of entry
  - 51 same col.)w/ 54 same col.)w/o
  - 52 same page)hdg. 55 same page)hdg.
  - 53 other) 56 other)
- 60 Non-consol. of heading only
  - 61 same col.)w/ 64 same col.)w/o
  - 62 same page)hdg. 65 same page)hdg.
  - 63 other) 66 other)
- 70 Non-consol. of subj. heading:
  - --counting subj. hdg. space
    - 71 same col.)subj. 77 same col.)whole
    - 72 same page)hdg. 78 same page)entry
    - 73 other)non-con. 79 other)non-con.
    - 74 same col.)entry
    - 75 same page)hdg.
    - 76 other)non-con.
  - -- not counting subject hdg. space
    - 7A same col.)whole 7P same col.)entry
    - 7B same page)entry 7Q same page) hdg.
    - 7C other)misfile 7R other)non-con.
    - 7E same col.)body 7X same col.)entry
    - 7F same page)only 7Y same page)non-
    - 7N other)misfile 7Z other)con.
- 80 Appearance only
- 90 Other or unknown

IV. Cause
- 10 Keying error
- 20 Variant cataloging practice
- 30 Program processing
- 40 Processing (not keying or prog.)
- 50 Inadequate design
- 60 Other
- 70 Unknown

V. Language
- 10 English
- 20 German
- 30 French
- 40 Spanish
- 50 Italian
- 60 Latin
- 70 Other Roman alphabet
- 80 Transliterated
- 90 Other

VI. Non-monographic type
- 10 Monographic series
- 20 Serial
- 30 Music score, map, other

21

FIGURE 1. FINAL VERSION OF ERROR CODE CATEGORIES

COL/Pg.	Col	Entry Number(s)	Type	Loc.	Effect	Cause	Lang.	Serial	Comments
06/702	1	01	20	11	63	20	10	—	- (B)
		01	35	44	80	30	10	—	-
		01	37	19	80	10	10	—	(x2)
		03	20	24	30	10	10	—	—
		05	50	21	61	20	10	—	—
		05	37	25	90	70	10	—	—
		09	20	23	30	30	10	—	—
		15	32	20	51	20	10	—	-14-
		18	50	20	54	30	10	—	8C 16 -
		19	32	23	30	20	10	—	—
		21	36	24	51	20	10	—	—
		21	33	25	80	70	10	—	—
		25	32	21	51	10	10	—	—
		27	70	20	51	20	10	—	8C 43 -
		33	70	20	40	40	10	20	—
		34	70	20	40	40	10	20	—
	2	01	70	10	40	40	10	20	—
		02	35	10	40	40	10	20	
		04	30	14	54	10	10	—	—
		06	70	10	40	40	10	20	—
		07	70	10	40	40	10	20	— 5C
		08	70	10	40	40	10	20	—
		09	70	10	40	40	10	20	—
		10	70	10	40	40	10	20	—
		11	70	10	40	40	10	20	—
		07	20	24	51	26	10	—	—
		11	60	26	40	30	10	—	—
		12	60	26	40	30	10	—	—
	3	12	32	24	51	10	10	—	—
		13	60	26	40	30	10	—	— 8C

Figure 2. Error Coding Data Sheet
13

The following section discusses the way the errors were finally categorized and coded.

C. ERROR CATEGORIZATION

Figure 1 shows the error categories used in this study and their numeric or alpha-numeric codes. Figure 2 shows an example of an error coding data sheet. The errors are recorded according to six general aspects: Type, Location, Effect, Cause, Language, and Non-Monographic Type. With one exception, each error was assigned only one code in each of the six aspects. Five of the aspects were suggested by Douglas Ferguson in his design paper.[11] The particular elements within this general categorization of error aspects have gone through at least 15 revisions during the course of our study.

Our goal was to produce a structure which would allow the categorization and recording of UCLCS errors in a useful, thorough, and unambiguous manner. This goal necessitated these six aspects and their detailed contents. Although we have no doubt that improvements could still be made in our error category structure, we believe the one presented here meets our goal and could also be readily modified to suit the purposes of other similar studies.

Following is a description of the first aspect, Type. Some discussion of the Effect and Cause aspects will be given at the same time, since these three aspects are closely related and difficult to explain in isolation. Further description of Effect and Cause and of the other three aspects follows.

1. Type

Aspect I is Type. One of the Types from 10 through 80 (see Error Code sheet, Figure 1) was selected as best describing each error encountered. It was nearly always possible to place an error within one of the specific categories (10-70), but occasionally an error did not fit into any of these and was therefore coded Type 80.

Type 10 (Duplicate data not suppressed) was used when an entry should have been consolidated with another entry and there is no apparent reason for the failure to consolidate. Generally, some discrepancy between two unconsolidated entries can be seen (for example, a typographical error or some difference in the way the two entries were cataloged), but occasionally the entries appear totally identical. The two entries in the example below are identical and should have consolidated, so there is no difference in the data as the entries appear on the page.

KING, Edward Louis, 1920-
—How chemical reactions occur, an introduction to chemical kinetics and reaction mechanisms. New York, W. A. Benjamin, 1963. 148 p. illus. 22 cm. (The General chemistry monograph series) Includes bibliography. 1. Chemical reaction, Rate of. QD501.K7514 541.39 63-8158 1126414
QD 501 K576h—L Chemistry Library;
QD501 K612—B Biochem. Library;
QD501 K55—D QD501 K54—R PHYS. SCI.;
QD 501 K44—SB; QD 501 K52—SD;
QD 501 K52—SD S & E.

—How chemical reactions occur, an introduction to chemical kinetics and reaction mechanisms. New York, W. A. Benjamin, 1963. 148 p. illus. 22 cm. (The General chemistry monograph series) Includes bibliography. 1. Chemical reaction, Rate of. QD501.K7514 541.39 63-8158 1296744
QD501 K52h 1963—SF; QD501.7514—SC

23

Figure 3. Example of Error Type 10--Duplicate Data Not Suppressed

The fact that these entries should have consolidated is reflected in the Effect category and the apparent (imputed) reason for this (a failure of the consolidation program in this case) is recorded in aspect IV-Cause.

Type 20 (Variant form of same data or edition) was used when two or more entries represented the same edition of a work but were cataloged differently by the different campuses. (It might be argued that this is not really an error in a technical sense.) Generally, in order for two works to be considered the same edition, they had to have the same author, title, publication date and collation statement. (The precise operational definition of edition as used in designing the programs and in this study is shown in Appendix C.) This error type may also occur when two entries should have consolidated; it may also be associated with a misfile. These factors are recorded in the Effect column. Generally this type of error was caused by variant cataloging practice, coded as Cause 20. An example of error Type 20 is shown below in Figure 4.

OSTERHELD, Dora Miller.
—Reference syllabus for use in advanced reference classes. 1965. *See* CAVANAGH, Gladys.
Z 1035 C314r 1965—L

OSTERHELD, Dora Miller, joint author.
—Reference syllabus, for use in advanced reference classes. 1965. *See* WISCONSIN UNIVERSITY. SCHOOL OF LIBRARY SCIENCE.
R Z1035 W57 1965—R

—Reference syllabus for use in advanced reference classes. 1965. *See* Wisconsin (State) University. Library School.
Z711 W56 1965—B Library School

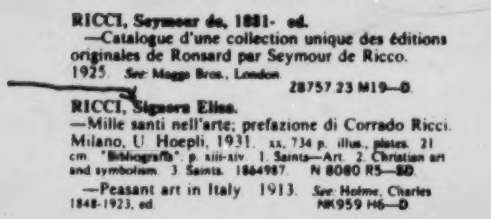
—Reference syllabus for use in advanced reference classes. 1965. *See* Wisconsin. University. Library School.
Z1035 W864 1965—D

Figure 4. Example of Error Type 20--Variant Cataloging Practice

The Type 30s include all sorts of typographical errors and are usually, but not always, caused by a keying error (Cause 10). Program processing (Cause 30) may also result in some of the Type 30s. These error types are fairly self explanatory. Types 32 and 33 (missing string and added string) include missing or added punctuation but exclude missing or added blanks. Type 37 includes all instances of orthographic inaccuracy not covered by the earlier listed categories. The most typical instances of Type 37 were misplaced or incorrect accent marks and the replacement of a correct letter by an incorrect letter. Note the example below which includes a number of types of orthographic inaccuracies.

Figure 5. Example of Error Type 30--Orthographic Inaccuracy

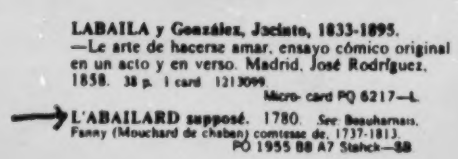
Type 40 (String improperly used in filing) is also self explanatory. This category includes not only instances of improper filing of articles (Types 44 and 45) but also cases where function terms, such as "jt. auth.," "ed.," "comp.," or associated titles, such as "Sir" or "1st Baron," were used in determining filing sequence. In UCUCS, foreign language initial articles, all function terms and associated titles were considered in filing, frequently resulting in misfiling of such entries. The examples below illustrate some of the Type 40 categories.



RICCI, Seymour de, 1881- ed.
—Catalogue d'une collection unique des éditions originales de Ronsard par Seymour de Ricco. 1925. *See* Maggs Bros., London
Z8757.23 M19—D

RICCI, Signora Elisa.
—Mille santi nell'arte; prefazione di Corrado Ricci. Milano, U. Hoepli, 1931. xx, 734 p. illus., plates. 21 cm. "Bibliografia", p. xiii-xiv. 1. Saints—Art. 2. Christian art and symbolism. 3. Saints. 1864987. N 8080 R5—SB.
—Peasant art in Italy. 1913. *See* Holme, Charles 1848-1923, ed.
NK959 H6—D

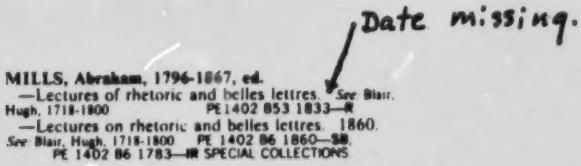
Figure 6. Example of Error Type 43--Misfile on Associated Title



LABAILA y González, Jacinto, 1833-1895.
—Le arte de hacerse amar, ensayo cómico original en un acto y en verso. Madrid, José Rodríguez, 1858. 38 p. 1 card. 1213099
Micro-card PQ 6217—L.
L'ABAILARD supposé. 1780. *See* Beauharnais, Fanny (Mouchard de chaban) comtesse de, 1737-1813.
PQ 1955 B8 A7 Stalick—SB

Figure 7. Example of Error Type 45 --Misfile on Non-English Article

Occasionally an entire data field (such as the date, collation, or title) would be missing from the entry. In this case, Type 50 was coded. Note the example below, in which the date is missing. Problems such as the example below could also be due to variant cataloging rules, but they were still coded as "data field missing." The specific example below also has a failure to consolidate because of spelling variations (of, on).



MILLS, Abraham, 1796-1867, ed.
—Lectures of rhetoric and belles lettres. *See* Blair, Hugh, 1718-1800 PE1402 B53 1833—R
—Lectures on rhetoric and belles lettres. 1860. *See* Blair, Hugh, 1718-1800 PE 1402 B6 1860—SB.
PE 1402 B6 1783—R SPECIAL COLLECTIONS

Figure 8. Example of Error Type 50--Data Field Missing

Conversely, sometimes data fields appeared where they should not have appeared. This usually occurred in subject or other added entries. Added entries were intended to include only the following data elements: author (if present), short title, date, campus location code and call number. Therefore if a subtitle, publisher, or edition statement appeared in an added entry, it was recorded as an error of Type 60. This is not so serious an error as many other types, but it does frequently result in wasted catalog space (Effect 40). An error was determined to have the effect of wasted

catalog space only when it resulted in at least one additional line of space being used. Notice the example in Figure 9. The first entry has two instances of Type 60, the first with Effect 90 (Other or unknown) and the second with Effect 40 (Wasted space).

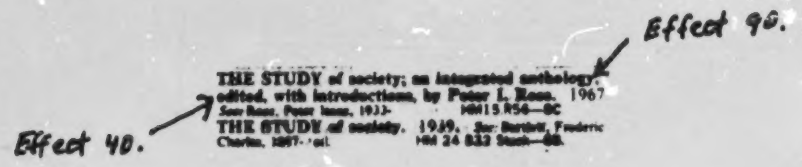

Effect 90.
THE STUDY of society; an integrated anthology.
edited, with introductions, by Peter L. Ross. 1967
THE STUDY of society. 1949.
Effect 40.

Figure 9. Examples of Error Type 60--Data Field Added; Effect 40--Wasted space

When a Type 60 did not result in at least one line being unnecessarily used, the Effect category was coded as 90 (Other or unknown). Because the programs were intended to prevent these data fields from being printed but failed to do so in these cases, the Cause category is coded Cause 30 (Program processing). However, it was later recognized that this error was more likely caused by a program design limitation. The problem occurred because the automatic format recognition program was not designed to be 100% perfect in delimiting $b subfield for subtitle, and there was no human post-edit to catch the deviations.

UCUCS was expected to include only monographic materials; serials, music scores, phonodiscs, maps, and other non-monographic materials were to be excluded. Therefore, an instance of inclusion of such non-monographic material is in some sense an error or deviation from intent. Type 70 was coded whenever non-monographic material was encountered. This is an error not only because the catalog was intended to include only monographic materials but also because failure to exclude non-monographs wastes space in the catalog and therefore costs additional money and user time. All such cases were coded: Type 70, Effect 40, Cause 40. Cause 40 includes all processing which is not included in the keyboarding process or in program processing. In the case of inclusion of non-monographic materials, such entries should have been but were not excluded at some point in the manual procedure of selecting records to be sent to ILR or in selecting those to be microfilmed and forwarded to the keying vendor. An example of the most typical sort of Type 70, a serial entry, is shown on the following page.

QUELLEN und Darstellungen zur hansischen Geschichte. n.F. [ser. 2] Bd.6- 1928- Weimar.
—Continues Hansische Geschichtsquellen. "Hrsg. vom Hansischen Geschichtsverein." Suspended 1. Hansa towns. 1099730.
DD 801 H18V5—4. Gross Collection Bus. Adm. Lib.
QUELLEN und Darstellungen zur Zeitgeschichte.
—Bd. 1- Stuttgart, Deutsche Verlags-Anstalt, 1957- v. 1. Germany—History—Sources. 1737129.
DD234 Q6—D.

Figure 10. Examples of Error Type 70--Inappropriate Entry

Type 80 (Other) included all error types which failed to fit into any of the above categories. This category was used rarely, usually for incorrect type size as in the example below.

MEETING on Automation in the Library; When, Where, and How, Purdue University, 1964. Papers. Edited by Theodore Andrews, with the assistance of Mirlete Morelock.Lafayette, Ind., Purdue University, 1965. 89 p.
Meeting held October 2-3, 1964. "Sponsored by the Purdue University Libraries with the cooperation of the Indiana Chapter of the Special Libraries Association and the Indiana Chapter of the American Documentation Institute." 1. Libraries —Automation. I. Andrews, Theodore, ed. II. Purdue University, Lafayette, Ind. Libraries. T. Automation in the Library. 1573714.
Z 678.9 M4 1964—BB; Z699 M4—B

Figure 11. Example of Error Type 80--Other

2. Location

The second aspect by which each error was coded is II-Location. Errors occurring in a main entry were coded in one of the 10s; errors occuring in added entries in the Author/Title catalog were coded in the 20s; and errors occurring in the Subject catalog were coded in the 30s. Nearly always an error could be located precisely in a specific part of the entry--for example, in the short title of a main entry or in the see reference of an added entry. Sometimes, however, an error could not be located specifically. For example, when an instance of duplicate data not suppressed (Type 10) occurs and two or more main entries are not consolidated (Effect 50), then it is impossible to say that the error occurred in the entry heading, the title statement, edition statement, etc., but merely that it occurred in a main entry. In such cases, the Location code 10 was used. The same approach was used in analogous cases for added and subject entries.

This particular aspect is straightforward and unproblematic with one exception. Generally, it was possible to reflect in one numeric (or alpha-numeric) code where the error was located; "where the error is located" meant both "where it appears" and "where its source is" since normally the two types of "location" coincide. For example, a typographic error appearing in an edition statement (Location 13) is noticed there by the reader and actually occurs there in the machine readable form of the record. Since the added entries and subject entries are generated from the source record which appears in full as the main entry, those errors appearing in an added or subject entry are usually traceable to an error in the main entry. This usually presents no problem in coding, however. Errors appearing in added or subject entries are coded in the 20s or 30s respectively in order to determine the error rates in these types of entries. For example, an error below.

RHODES, David H., joint author.
—Anesthesia in clinical ophthalmology. 1963.
See Duncalf, Deryck. WO 200 D912a 1963—L;
RE82 D8—B Optom. Library; WO200 D8—D.
—Anesthesia in clinical ophthalmology. 1963.
See Duncalf, Deryck. RE82 D91a 1963—BF.

Figure 12. Example of Error in Which the Source of Probable Error Is Not Apparent from Sample Entries

One can only determine from these entries that they may be instances of the same edition of a work, and therefore it may be that they should have consolidated. The main entries must be checked to determine whether they are duplicates. Their respective main entries are shown below.

DUNCALF, Deryck.
—Anesthesia in clinical ophthalmology [by] Deryck Duncalf [and] David H. Rhodes. Baltimore, Williams & Wilkins, 1963. xviii, 168 p. illus. 24 cm. Bibliography: p. 153-163. 1. Anesthesia in ophthalmology. I. Rhodes, David H., joint author. RE82.D8 617.96 63-20432 Biomedical Library I. Anesthesia-Ophthalmology II. Title 1136517
WO 200 D912a 1963—L;
RE82 D8—B Optom. Library; WO200 D8—D.
—Anesthesia in clinical ophthalmology. [by] Deryck Duncalf [and] David H. Rhodes. Baltimore, Williams - Wilkins, 1963. xviii, 168 p. illus. 1. Anesthesia. 2. Ophthalmology. I. Rhodes, David H. 1803170. WW 168 D911a 1963—BD.
—Anesthesia in clinical ophthalmology [by] Deryck Duncalf [and] David H. Rhodes. Baltimore, Williams & Wilkins, 1963. xviii, 168 p. illus. 24 cm. Bibliography: p. 153-163. 1. Anesthesia in ophthalmology. I. Rhodes, David H., joint author. RE82.D8 617.96 63-20432 1306717 RE82 D91a 1963—BF.

Figure 13. Main Entry Form of Sample Added Entries Shown in Figure 12

Here we see that the entries do represent the same edition and therefore should have consolidated but did not. The question then arises whether to indicate in the Location code where the error appears or where it is "caused:" whether to indicate Location 20 or Location 17. The authors wanted very

much to tabulate data both according to the location of error appearance and to the location of error source. We wanted to be able to say what percentage of the errors appeared in added and subject entries, what percentage of the added and subject entries contained at least one error, etc. But we also wanted to trace an error to its source location where possible, primarily in order to identify which parts of the computer programs could best be improved. The dilemma was resolved by breaking the general rule of "each error is to be assigned only one code for each of the six general categories." Both the source location (i.e., a main entry code in the 10s) and the location of error appearance (either 20 or 30) were recorded in the recorded column. Later, when the data was keypunched, a separate column was created to handle this situation, and the data reduction program was designed to tabulate either of the two columns needed for a specific purpose.

As with the main entries, when an error could not be traced to its precise location either in an added or subject entry or in its corresponding main entry, the general category (20 or 30) was entered in the Location code. In cases where an extra data field was added to an added or subject entry, the location of the error was recorded as 26 or 36, respectively.

3. Effect

The third aspect according to which each error was coded is III—Effect. Each error has at least one effect, sometimes more than one. In each case, however, only one effect was assigned to each error. This rule occasionally presented some conflict, but the conflicts were generally reconcilable by fairly rational means.

The first of the Effects, 10 (Lost entry point), is self-explanatory. It was used when there would be no way to find an item by means of an entry point which should have been available. Instances of misfiling where the entry is filed far from the proper filing point could be considered instances of lost entry points. However, in our study such errors were coded under Effect 20 (Misfiled). Misfiles which are so drastic as to have the effect of a lost entry point were grouped with the Effect 10s later in data reduction as being instances of "Fatal" errors.

Effect 10 was used relatively rarely in this study because one planned section of the study was not carried out. It was originally intended that a sample of the microfilmed source records to be taken and that all of the appropriate entry points in these records be looked up in UCUCS. Another less thorough way to check for lost entry points would be to look up the appropriate added and subject entries for each main entry in our Author/Title sample. This step was not undertaken either. A few lost entry points were noticed in the course of our study, however, and we believe that a systematic effort to measure this source of error should be undertaken.

Effect 20 (Misfiled) includes all entries which have misfiled but in which the misfile is not due to failure to consolidate. All misfiled entries were categorized according to whether they were misfiled but were within their appropriate column (21 and 24), misfiled into another column on the same page (22 and 25), or misfiled onto another page (23 and 26). Each of these categories was further divided into cases in which the entire entry (heading plus body of the entry) misfiled, and those in which only the body of the entry (without the entry heading) misfiled. An example of "same column" misfile in which the entire entry has misfiled is shown below in Figure 14.

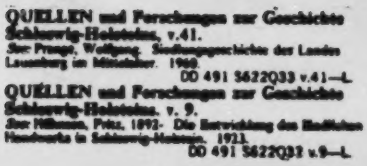

QUELLEN und Forschungen zur Geschichte Schleswig-Holsteins, v.41.
Der Prange, Wolfgang. Siedlungsgeschichte der Landes Lauenburg im Mittelalter. 1968.
DD 491 S622Q33 v.41—L
QUELLEN und Forschungen zur Geschichte Schleswig-Holsteins, v. 9.
Der Hansen, Peter. 1892- Die Entwicklung des ländlichen Handwerks in Schleswig-Holstein. 1923.
DD 491 S622Q33 v.9—L

Figure 14. Example of Effect 21—Entire Entry Misfiled in Same Column

An example in which an entry without its heading misfiled in the same column is shown below in Figure 15.

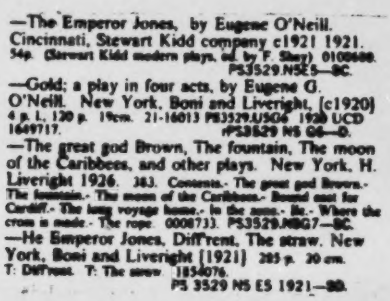
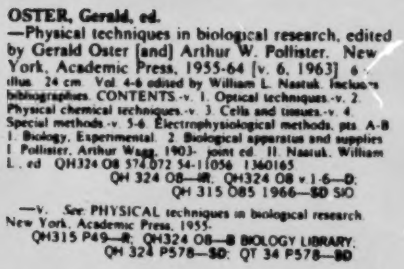

—The Emperor Jones, by Eugene O'Neill. Cincinnati, Stewart Kidd company c1921 1921. 54p. (Stewart Kidd modern plays, ed. by F. Shay) 0100600. PS3529.N5E5—BC.
—Gold; a play in four acts, by Eugene G. O'Neill. New York, Boni and Liveright, [c1920] 4 p. l., 120 p. 19cm. 21-16013 PS3529.U5G6 1920 UCD 1649717. rPS3529 N5 G6—G.
—The great god Brown, The fountain, The moon of the Caribbees, and other plays. New York, H. Liveright 1926. 383. Contents.- The great god Brown.- The fountain.- The moon of the Caribbees.- Bound east for Cardiff.- The long voyage home.- In the zone.- Ile.- Where the cross is made.- The rope. 0008733. PS3529.N8G7—BC.
—He Emperor Jones, Diff'rent, The straw. New York, Boni and Liveright [1921] 283 p. 20 cm. T: Diff'rent. T: The straw 1854076. PS 3529 N5 E5 1921—BD.

Figure 15. Example of Effect 24--Entry Without Heading Misfiled in Same Column

Effect 30 (Content uncertain) was used whenever an error resulted in some confusion or doubt about the exact meaning of some part of the entry. Frequently this category was used when a typographical error left the meaning of a word or phrase less than certain. "Uncertain" was strictly interpreted by the authors. There were many cases where a typographical error, for example, might have been interpreted by others as merely affecting the appearance of the word or data element because they would have been able to guess with some degree of confidence what was meant. We used Category 30 instead of 80, however, when there was any reasonable chance that the error would have been confusing to the catalog user. Since the "catalog user" was defined as being non-librarians as well as librarians, we tended not to give the benefit of the doubt to the catalog in these cases. A typical example is shown in Figure 16. (The second entry is probably due to an AFR error.)



OSTER, Gerald, ed.
—Physical techniques in biological research, edited by Gerald Oster [and] Arthur W. Pollister. New York, Academic Press, 1955-64 [v. 6, 1963] 6 v. illus. 24 cm. Vol. 4-6 edited by William L. Nastuk. Includes bibliographies. CONTENTS.-v. 1. Optical techniques.-v. 2. Physical chemical techniques.-v. 3. Cells and tissues.-v. 4. Special methods.-v. 5-6. Electrophysiological methods, pts. A-B
1. Biology, Experimental. 2. Biological apparatus and supplies
I. Pollister, Arthur Wasa, 1903- joint ed. II. Nastuk, William L., ed. QH324.O8 574.072 54-11056 1360165
QH 324 O8—R; QH324 O8 v.1-6—D;
QH 315 O85 1966—BD SIO
—V. See: PHYSICAL techniques in biological research.
New York, Academic Press, 1955-
QH315 P49—R; QH324 O8—B BIOLOGY LIBRARY;
QH 324 P578—BD; QT 34 P578—BD

Figure 16. Example of Effect 30--Content Uncertain

Effect 40 (Wasted space--not non-consolidation) was discussed earlier. It was used whenever an error resulted in the space of at least one additional line being used in an entry. This frequently occurred in conjunction with an added data element (Type 60) and always occurred when a non-monographic entry was included in the catalog (Type 70). Although errors resulting in non-consolidation also result in space being wasted, they were coded under one of the non-consolidation categories (50 through 7Z) explained later in this section. In measuring the amount of space wasted in the catalog due to all kinds of errors, the 40 category is to be used in conjunction with the 50-7Z categories to arrive at an estimate.

Care was taken in designing the error code structure to be able to estimate the seriousness of the non-consolidation problems in UCUCS. This goal necessitated a very specific and rather complex breakdown of types of non-consolidation. Non-consolidation can occur in main, added, or subject entries; the type of entry which has failed to consolidate will affect the amount of space wasted by that failure. Also, the entire entry may fail to consolidate, or just its heading, or, in the case of subject entries, the subject heading may fail to consolidate, and this in turn may result in failure of entire entries or just entry headings not to consolidate.

It was assumed that any non-consolidation also represents a deviation of sorts, even in cases where the entries are adjacent. In the following paragraphs we will attempt to state the meaning of each of the non-consolidation categories.

Effect 50 (Non-consolidation of entry) includes several subdivisions. One of the 50s Effect Codes is used when the body of the entry should have consolidated with another entry but failed to do so. Codes 51, 52 and 53 are used when the entire entry (body plus heading) failed to consolidate. (Code 51 is used when the entry should have consolidated with another entry in the same column, 52 when it should have consolidated with another entry on the same page, and 53 when it should have consolidated with another entry on another page.) Codes 54, 55, and 56 are used when an entry without a heading should have consolidated; 54 is for the same column, 55 for the same page, and 56 for another page. Figure 17 illustrates the use of code 51 (entire entry non-consolidation including heading, same column) and Figure 18 shows three instances of 54 (only body of entry failed to consolidate, same column).

OSTEOTOMY at the upper end of the femur.
1965. *See* Milch, Henry, 1895-1964.
RD560 M630 1965—BF.
OSTEOTOMY at the upper end of the femur.
1965. *See* MILCH, Henry, 1895-1964.
WE 885 M639o 1965—SD.
OSTEOTOMY at the upper end of the femur.
1965. *See* Milch, Henry, 1895-1964.
WE 885 M591o 1965—L.

Figure 17. Examples of Effect 51--Entire Entry Non-Consolidation, Same Column

OSTER, Daniel, ed.
—aŭvres complètes. 1964. *See* Montesquieu, Charles Louis de Secondat, baron de la Brède et de, 1689-1755.
PQ2011 A1 1964—B.
—Œuvres complètes. 1964. *See* Montesquieu, Charles Louis de Secondat, baron de la Brède et de, 1689-1755.
PQ2011 A1 1964—B; PQ 2011 A1 1964—BB.
—Oeuvres complètes. 1964. *See* Montesquieu, Charles Louis de Secondat, baron de la Brède et de, 1689-1755.
PQ 2011 A1 1964—BD.
—OEuvres complètes. Pref. de Georges Vedel. Presentation et notes de Daniel Oster. New York, Macmillan 1964. 1117 illus. Bibliographical footnotes.
I. Montesquieu, Charles Louis de Secondat, baron de la Brè
II. de et de, 1689-1755. 0003365
PQ2011 A1 1964—BC.

Figure 18. Examples of Effect 54--Only Body of Entry Non-Consolidation, Same Column

Effect 60 (Non-consolidation of heading only) has the same subdivisions as Effect 50. The 60 codes are used when the body of the entry did not consolidate correctly (i.e., the entry represents a different work or edition from any other on that sample page), but the entry's heading should have consolidated and failed to do so. Even if the entry associated with a particular non-consolidated heading should not have consolidated, it will still be misfiled as a result of its heading failing to consolidate. Therefore, the 60s codes indicate whether the entry is misfiled onto the same column, same page, or another page from where it should be filed. The first three (61, 62, and 63) are used for entries with headings "attached" which have been misfiled because of non-consolidation of headings. See Figure 19 for an illustration of a 61. The second three (64, 65, and 66) are used for entries which consist only of an entry body, but whose headings (actually appearing with an entry filed ahead of them) failed to consolidate. Again an example is needed to make this intelligible; Figure 20 shows an instance of a 64.

KING, Edward Jasper, 1916-
—Acid-base equilibria, by Edward J. King. [1st ed.] oxford, New York, Pergamon Press [published in the Western Hemisphere by Macmillan, New York [1965] xi, 341 p. illus. 24 cm. (The International encyclopedia of physical chemistry and chemical physics. Topic 15: Equilibrium properties of electrolyte solutions, v. 4) Includes bibliographies. A Pergamon Press book. 1. Acid-base equilibrium. 3. The International encyclopedia of physical chemistry and chemical physics (Oxford) Topic 15: Equilibrium properties of electrolyte solutions, v. 4. QD453.I 5 topic 15, vol. 4 541.392 65-2709
1567560.
QD 453 I5 v.15:4—BB; QD453 I5 ser.15 v.4—B;
QD 501 K575e—L Chemistry Library;
R QD453 I58 topic 15 v.4—R Phys.Sci.;
R QD453 I58 topic 15, v.4—R Phys. Sci.;
QD 453 I5 v.15:4—BB; QD 453 I5 v.15:4—BB;
QD 501 K52—BB 5 I O.
—Qualitative analysis and electrolytic solutions. Under the general editorship of Larkin H. Farinholt. New York, Harcourt, Brace [1959] 641 p. illus. 24 cm. Includes bibliography. 1. Chemistry, Analytic —Qualitative. 2. Electrolytes. 3. Solution (Chemistry).
QD81.K46 540.1 59-7733 aCD 1628204
QD81 K46—B; QD 81 K46—BB; QD 81 K5—BB;
QD 81 K52—BB 5 & L.
KING, Edward Jasper, 1916- joint author.
—General chemistry. 1967. *See* Pryde, Marvin Ambrose, 1910-
QD 33 P9—BB.

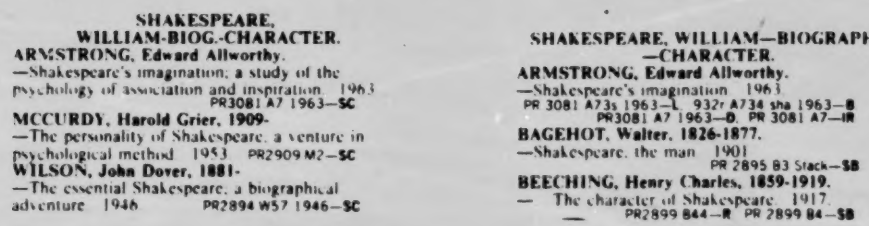
Figure 19. Example of Effect 61--Non-Consolidation of Heading Only, Resulting in Misfile of Entire Entry, Same Column

KING, Ernest Joseph, 1878-
—...United States navy at war, 1945. *See* U. S. Office of naval operations.
D 773 A5 1945—BB.
—The War reports of General of the Army George C. Marshall, Chief of Staff, General of the Army H. H. Arnold, Commanding General, Army Air Forces. 1947. *See* Title.
D 769 W3—BB.
KING, Ernest Joseph, 1878-1956.
—Fleet Admiral King, a naval record by Ernest J. King and Walter Muir Whitehill. [1st ed.] New York, W. W. Norton [1952] xv, 674 p. illus., ports., maps. 24 cm. 1. United States. Navy—History. 2. World War, 1939-1945—Naval operations, American. I. Whitehill, Walter Muir, 1905- joint author. E182.K53 923.573 53-13493
UCR 1389111. E182 K56—B; E 182 K5—BB.

Figure 20. Example of Effect 64--Non-Consolidation of Entry Heading Only, Resulting in Misfile of Body of Entry Only, Same Column

It may be noted that the lack of interfiling of main and added entries for the same name is actually a filing policy or practice, not an "error." We can say it is, in UCUCS, a deviation from a desired practice, but the desired practice was not implemented because, by management decision, the data was not encoded to accomplish interfiling.

Effect 70 (Non-consolidation of subject heading) contains many subdivisions. The complexities of the above two general non-consolidation categories (50s and 60s) pale by comparison to the 70s category. This category combines all the previous possibilities with the non-consolidation of an entry's subject heading. There are 22 possible codes when a subject heading has failed to consolidate. These codes are first divided into two groups: those in which the space taken up by the subject heading itself should be included in this particular error coding, and those in which the space taken up by the subject heading should not be included in the code. To make clear why this division is necessary, note Figure 21 below, in which the cause of the problem is the failure of the authority control software.



SHAKESPEARE,
WILLIAM-BIOG.-CHARACTER.
ARMSTRONG, Edward Allworthy.
—Shakespeare's imagination; a study of the psychology of association and inspiration. 1963
PR3081 A7 1963—SC
MCCURDY, Harold Grier, 1909-
—The personality of Shakespeare, a venture in psychological method. 1953. PR2909 M2—SC
WILSON, John Dover, 1881-
—The essential Shakespeare, a biographical adventure. 1946. PR2894 W57 1946—SC

SHAKESPEARE, WILLIAM—BIOGRAPHY
—CHARACTER.
ARMSTRONG, Edward Allworthy.
—Shakespeare's imagination. 1963.
PR 3081 A73s 1963—L 932r A734 sha 1963—B
PR3081 A7 1963—D PR 3081 A7—IR
BAGEHOT, Walter, 1826-1877.
—Shakespeare, the man. 1901.
PR 2895 B3 Stack—SB
BEECHING, Henry Charles, 1859-1919.
— The character of Shakespeare. 1917.
PR2899 B44—R PR 2899 B4—SB

Figure 21. Example of Subject Heading Non-Consolidation

In this illustration, the first subject heading failed to consolidate with the second subject heading which appeared in slightly different form. This fact is reflected in coding each of three entries appearing under the first subject heading. But the amount of space wasted by the non-consolidated subject heading should be tallied only once. In such cases, then, the first entry listed under a non-consolidated subject heading was coded to include consideration of the amount of space wasted by the subject heading, and subsequent entries under the same subject heading were tallied so as not to include the wasted space of the subject heading.

In each instance, after deciding whether the subject heading space should be considered in coding an entry in error, the next step is analogous to determining the appropriate 50s or 60s code as discussed previously. One needs to determine exactly what part(s) of the entry (if any) have failed to consolidate and whether the entry was misfiled on the same column, same page or another page. Codes 71, 72, and 73 are used when no part of the entry (except its subject heading) should have consolidated, so that it has misfiled only because its subject heading was inappropriately duplicated. Codes 74, 75, and 76 are used when not only the entry's subject heading but also the entry heading failed to consolidate, but the body of the entry correctly did not consolidate. Codes 77, 78, and 79 are used

when not only the subject and entry heading but also the body of the entry should have consolidated. The nine codes listed above are only used when the space wasted by the non-consolidated subject heading is to be counted; that is, they are used only for the first entry appearing under a particular subject heading.

Entries in which the subject heading space is not to be considered--that is, second, third, and subsequent entries under a subject heading--are coded with one of the 12 alpha-numeric 70s codes. Here, in cases where only the subject heading itself was supposed to have consolidated (analogous to the 71, 72, and 73 above), a separate division must be made for cases in which the entire entry (heading plus body) misfiled as a result of the subject heading non-consolidation and for those cases in which only the body of an entry misfiled. Figure 22 shows an instance of an entire entry misfiling because of its subject heading failing to consolidate.

SHAKESPEARE,
WILLIAM—BIBLIOGRAPHY—FOLIOS.
1623.
SHAKESPEARE Association, London.
—Studies in the first folio, written for the Shakespeare Association in celebration of the first folio tercentenary and read at meetings of the Association held at King's College, University of London, May-June, 1923. 1924.
Z 8813 S5—SB

SHAKESPEARE, WILLIAM—BIBLIOGRAPHY
—FOLIOS. 1623.
COLE, George Watson, 1850-1939.
—The first folio of Shakespeare. 1909.
Z 1008 847p v 3—L

Figure 22. Example of Effect 7A—Subject Heading Non-Consolidation Resulting in Entire Entry Misfile, Same Column

The other six codes in this category are used when either the entry heading should have consolidated but did not (codes 7P, 7Q, 7R), or when the entire entry failed to consolidate (codes 7X, 7Y, and 7Z). An illustration of each type of case is shown in Figures 23 and 24.

SHAKESPEARE, WILLIAM-BIOG.
ADAMS, Joseph Quincy, 1881-
—A life of William Shakespeare, by ..... Library ed. 1925. PR2894 A3 1951—SC
ALEXANDER, Peter, 1893-
—A Shakespeare primer. 1961
PR2895 A4 1961—SC

SHAKESPEARE, WILLIAM—BIOGRAPHY.
ADAMS, Joseph Quincy, 1881-
—A life of William Shakespeare. 1951.
PR 2894 A3 1951—SB
ALEXANDER, Peter, 1893-
—A Shakespeare primer. 1951.
PR2895 A43 1951a—R

Figure 23. Example of Effect 7Q—Subject Heading Non-Consolidation Resulting in Entry Heading Non-consolidation, Same Page

EISENHOWER, DWIGHT DAVID, PRES. U. S., 1890-.
ARMY Times, Washington, D.C.
—The challenge and the triumph; the story of General Dwight D. Eisenhower, by the editors of the Army times. 1966. E836.A87—SC

EISENHOWER, DWIGHT DAVID, PRES. U. S., 1890-.
ADAMS, Sherman, 1899-
—Firsthand report. 1961.
E835 A33—R; E 835 A4 1961—SB
ALBERTSON, Dean, 1920-ed.
—Eisenhower as President. 1964
E 836 A8 1964—R E 836 A8 1964. E836 A65—R; E 836 A33e—L; E 836 A4—SB
AMBROSE, Stephen E.
—Eisenhower and Berlin, 1945. 1967.
D 755.7 A45—R; D755.7 A45—R; D 755.7 A5—SB
ARMY times, Washington, D. C.
—The challenge and the triumph. 1966.
E836 A72—R; E 836 A87—R

Figure 24. Example of Effect 7X—Subject Heading Non-Consolidation Resulting in Entire Entry Non-consolidation, Same Column

Category 80 was used when the only result of an error was some adverse effect on the appearance of the catalog. For example, it could have made the entry difficult to read or just unattractive. But if the error might conceivably have resulted in some confusion on the part of the user regarding the content of the entry, Code 30 was used instead of Code 80. An example of a number of instances of Effect 80 appears in Figure 25.

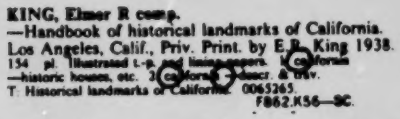

KING, Elmer R comp.
—Handbook of historical landmarks of California.
Los Angeles, Calif., Priv. Print. by E.R. King 1938.
154 pl. Illustrated t.-p. and lining papers. 1. California —historic houses, etc. 2. California—Descr. & trav.
T. Historical landmarks of California. 0065265.
F862.K56—SC.

Figure 25. Examples of Effect 80--Appearance Only

The 90 code for Other or unknown was rarely used. Its primary use was in conjunction with program processing errors which printed out unnecessary data fields but which resulted in no wasted space in the catalog. (Type 60, Effect 90, Cause 30.) Note the example in Figure 26.

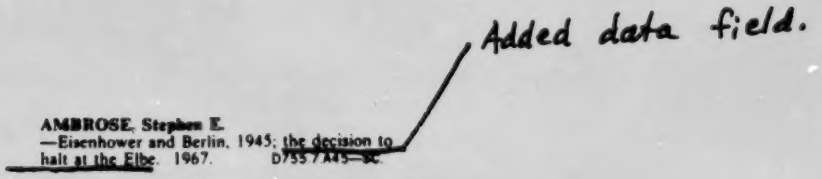


Figure 26. Example of Effect 90--Other or Unknown

4. Cause

The fourth aspect by which each error was coded is Cause. Whenever possible the cause of an error was determined and coded. It was rarely possible to determine with absolute certainty the cause of a particular error, although usually we could be fairly confident. When we suspected the validity of our opinion of an error's cause, we coded Cause 70 (Unknown). In general, though, we attempted to assign each error a specific cause.

Cause 10 (Keying error) is self explanatory. This code includes all those mistakes that we normally think of as typographical or keyboarding errors.

Cause 20 (Variant cataloging practice) includes all instances where a discrepancy in the way two or more campuses (or even libraries within the same campus) cataloged a given work resulted in some deviation or inconsistency

in the catalog. Frequently, some minor discrepancy would result in non-consolidation of records; occasionally variant cataloging practice resulted in misfiling of a record. Such cases were counted. But instances in which a discrepancy in cataloging happened to be noticed but did not result in error were not tallied.

Cause 30 (Program processing) includes all instances in which an error was caused by the failure of a program to do what it was designed to do. This could have been any of the various programs used to produce UCUCS: automatic format recognition, generation of added entries, the print program, or whatever.

Cause 40 (Processing other than keying or programming) was coded when there was apparently some slip-up in processing which was not program processing or keyboarding. This code was used primarily for instances of non-monographic material being inctuded in the catalog.

Cause 50 (Inadequate design) was used to cover cases in which a program should or could have been written to handle a particular type of situation. For example, when an English article is improperly used in filing, the error was coded Cause 30 (Program processing) because a program was written to prevent such occurrences; the program apparently failed in this particular case. But when a non-English article is improperly used in filing, we coded for Cause 50 (Inadequate design) since no program was written to suppress these articles in filing, and conceivably one could have been written to handle such situations, given sufficient time and budget.

Cause 60 (Other) was to be used when the cause did not fit any of the above categories but was known to us. It turned out that Cause 60 was only rarely assigned in this study. Cause 70 (Unknown) was used when we could not determine the cause of the error within the realm of reasonable doubt.

5. Language

The fifth aspect, V--Language, is self explanatory. The language in which the error itself was found was the language which was coded. That is, if an error occurred in a title of a work and the title was written in French, the code used was 30. But if the error occurred in a subject tracing which was written in English, the code used was 10, even though the title and the work itself may have been in French. In those cases discussed earlier in which the error cannot be pinpointed to a specific location in the entry, the Language code is used which reflects the language of the title of the work.

6. Non-Monographic Type

The final aspect is VI--Non-monographic type. The first item in this category, 10 (Monographic series), does not imply that monographic series should not be included in the catalog. They should have been and were included. The category is given here to evaluate errors which happen to occur in monographic series records. It was thought that different error patterns might emerge in such records. Thus no error was recorded just because an entry happened to be a monographic series, but if an error occurred in such a record it was coded 10 in this category.

As mentioned previously, only monographs were intended by the system designers to be included in this catalog. Serials, music scores, maps, phonodiscs, etc., were intended to be excluded, but such records were not always successfully omitted. When such a record was found, it was coded here either 20 or 30 depending on the type of entry. This category was used only to code non-monographic entries; for ordinary monographic entries a dash was placed in the Non-monographic type category.

7. Comments

The final column in the Error Coding Data Sheet was reserved for comments. It was used, especially in the early part of the study, for noting problems or questions which needed to be discussed and resolved by the two authors doing the detailed review. It was also used to record multiple instances of the same type of error. For example, if there were four typographic errors in the title of an entry with the same cause, effect, etc., the first instance of the error would be coded in the six categories and "X 4" would be noted in red in the comments column. In this way the keypuncher was alerted simply to punch one card and duplicate it three times.

The "Comments" column was also used to record that an entry was from the Santa Cruz campus only. The Santa Cruz records were sent to ILR in machine readable form, unlike the rest of the campuses which sent catalog cards. Since certain types of errors were observed by the authors to occur exclusively or nearly exclusively in Santa Cruz records, we decided to note all errors in Santa Cruz records by writing "SC" in the "Comments" column. As many of these errors may be due to the fact that the records were already in machine-readable form when ILR received them, it was felt that recording Santa Cruz record errors would facilitate improving the separate programs designed to reconcile these records with the rest of the file.

D. METHOD OF CONVERSION OF COLLECTED DATA INTO MACHINE READABLE FORM FOR TABULATION

The data were carefully collected on code sheets such as the one shown earlier in Figure 2, and then punched onto IBM cards in the order in which they appeared on the sheets. One card was used for each error. Each card, then, started with a two-digit volume number, a three-digit page number, a one-digit column number, and a two-digit entry number. (Numbers in each category which had fewer digits were right justified and zeros were placed ahead of them.) The resulting eight-digit number represented a unique identification number by which an error could be relocated in the catalog.

The codes for the six major error categories then appeared consecutively in card columns 9 through 18. Card column 30 was used to record that an error was located in a Santa Cruz entry (as the number "1"); column 30 was left blank when the record was not from Santa Cruz.

In the section above describing the second error aspect, location, it was mentioned that in some cases two Location codes had to be recorded. This occurred when the place where the error was noticed (in an added entry, for example) was different from the source of the error (in a part of the main entry which did not appear in the added entry--for example, the publisher statement). In coding such errors, both types of location were noted in the Location column of the data sheet. But since both Location codes could not be punched in the same column of an IBM card, a separate part of the IBM card was used for one of the Location codes. It was arbitrarily decided that the place where an error appeared (that is, Location code 20 or 30) would be punched in the regular Location code column and that the source of the error (that is, Location codes 1A through 19) would be punched in columns 35 and 36 of the IBM card. Accordingly, the data sheets were all re-scanned and multiple entries in the Location column were erased. The code 20 or 30 was written in the Location column, and the source location code was noted in the center of the Comments column.

One other modification of the recorded data was necessary before keypunching the data was possible. Since the program used to tabulate the data could manipulate data only in numeric form and since some of the codes used in the study were alpha-numeric, these had to be changed to numeric form before keypunching. Accordingly, the Location codes 1A, 1B, 1C, 1D, 1X, and 1Y were erased on the data sheets and replaced with 41, 42, 43, 44, 45, and 46 respectively. Alpha-numeric codes had also been used in the third error aspect, Effect. Changing the alpha-numeric codes to numeric form was a little more problematic here since nine major categories were used (numbered 10, 20, ...90), and we did not want to use the same first digit for categories which were conceptually unrelated. That is, we did not want to number instances of Effect 70 (Non-consolidation of subject heading) uring the first digit of another category (31, 32, etc., for example). The problem was resolved in the following manner: the data recorded as Effect 7A, 7B, etc., were re-coded as 70 in the Effect column of the data sheets, and a separate column was created in the Comments column. The data in this part of the Comments column was keypunched in columns 38 and 39 of the IBM cards. This data consisted of the digits 01 through 12. By way of illustration, data originally coded as 7A and 7Y in the Effect column of the data sheet were recorded there as 70 and in the Comments column as 01 and 11, respectively.

All keypunching--except approximately 100 cards keyed and proofread by one of the authors--was done by the U.C. Berkeley Computer Center keypunchers. The punched cards were key verified, except for the hundred or so which were manually proofread. That is, approximately 98.5% of the data cards were key verified.

E. SUGGESTED IMPROVEMENTS IN THE METHODOLOGY

Hindsight is usually better than foresight, and carrying out this project proved to be no exception to that general rule. Hindsight has dictated a number of suggestions for improving our methodology. These will be discussed here so that anyone attempting a similar study can incorporate them into their methodology.

A shortcoming of the present design is that there is no allowance made for recording whether an entry was a title main entry or corporate author main entry. After most of the sample sheets had been analyzed, we began to notice that certain kinds of errors seemed to appear more frequently in title main entries than in author main entries. But this is merely a subjective impression, and we did not have time to re-design the error categories and re-examine the sample pages in order to record the necessary information. It should not be necessary to designate a separate, i.e., seventh, major category in order to record such information as the errors are described on the data sheets. One way of including the information would be to circle or underline the entry in the Location column if the error occurred in a title or corporate author main entry.

In this study, Cause 60 (Other) was rarely used. Probably a combined category for "Other or Unknown" would suffice for most studies.

There was no attempt to estimate the number of lost entry points in the catalog. In a few cases the analysts stumbled upon such instances, but there was no systematic attempt to discover the probable rate of lost entry points. Such a systematic attempt might be made by taking a random sample of the source records for the catalog and simply looking up all the entry points in the catalog indicated by the record. Where this is not possible, just checking whether all the entry points indicated by the main entries in the sample data sheets are in the catalog would give some idea of the number or extent of missing entry points. Of course, if the main entry itself were omitted from the data base, both it and all its added entries would be lost, and there would be no way of recording such occurrences if the latter method is used. For this reason, checking from source records is obviously a better method.

A fourth suggestion is to employ multi-lingual analysts if possible. One of the present authors has a reading knowledge of Spanish and Portuguese. Many other foreign language entries (152 of them) had to be read by other people familiar with these languages. This is time-consuming since one of the analysts had to go over the entries with these people to make sure the errors were coded consistently. If the authors had been able to read more of the language found in the catalog, this part of the study would have been accomplished more efficiently.

A very simple improvement would be to add a subcategory in the Type 30s (Orthographic inaccuracy) to include all instances of incorrect accent marks, umlauts, and other such characters. In the present study these errors were coded 37 (Other misspelling or undetermined), along with a wide variety of other kinds of errors.

Finally, an error in our procedure was the failure to have re-keyed (and then key verified) the hundred or so cards punched by one of the authors and manually proofread. There are some discrepanies in the tabulated results of the study which may be due to keying errors in this relatively small batch of punched cards which were not key verified. Some of the discrepancies (or possibly all of them) may result from other causes. Over 92,000 digits and letters were handwritten on the data sheets and then keypunched, so it is likely that some characters were illegibly written and misread by both the keypuncher and the key verifier. Moreover, errors could easily have resulted when the data sheets were rescanned and some of the data re-coded in order to suit the requirements of the data reduction program. Since nearly all the discrepancies are in the subject catalog section (where most of the re-coding was done), it seems likely that this factor contributed to the error. In any case, <u>all</u> the punched cards should have been key verified, not just 98.5% of them.

IV. RESULTS

A. INTRODUCTORY COMMENTS

Once the data had been keypunched and the computer program most appropriate to our needs had been identified by the programmer, it was relatively easy to generate tables correlating the various aspects of the errors recorded. We attempted, therefore, to produce all those tables which might be of interest to people associated with the UCUCS project or to people who might want to compare our results with those of similar future studies. The program was used to generate 48 tables, as well as other data such as the total number of errors in the Author/Title catalog and in the Subject catalog. All of the data produced by these programs are included in this report. Some of the data is introduced and discussed in this section of the report, namely those tables and results which seem most likely to be of general interest.

It was noticed by the authors early in the study that when one error was found in an entry, chances were that another would be found in the same entry. That is, many entries had no error, and it seemed that those which contained one error often had more than one. This was an interesting subjective observation, and it was therefore hoped that the program used in data reduction would be capable of determining the average number of errors per entry of the entries in error. The page, volume, column, and entry numbers of each error (together representing an entry identification number) were recorded, so theoretically it would have been possible for the program to store this information for each error, to record the number of errors associated with each of the entries which contained some error, and then to find the average number of errors per entry of the entries in error. But this feat was beyond the ken of the data reduction program used for this study.

As mentioned earlier in the study, it was noticed during the process of data collection that certain error patterns seemed to appear in records with a location code indicating they came from the Santa Cruz campus, the only campus which sent records already in machine-readable form and which had a different processing procedure from the remainder of the records. Therefore, errors found in Santa Cruz records were so noted on the data sheets. It was possible, then, to generate tables correlating any error aspects for errors occurring in Santa Cruz records just as it was possible to generate such tables for the entire body of data. All tables produced were therefore done for the entire body of data and also for the subset of data from the Santa Cruz campus.

In addition to the tabulation of error data peculiar to Santa Cruz entries, all of the tables presented in this report are generated for both the Author/Title catalog sample and for the Subject catalog sample. That is, each correlation of two error aspects--for example, error type and cause--appears in segregated tables for the two parts of the catalog. Moreover, as explained above, each correlation is also divided according to the data for the entire sample and for the Santa Cruz records only. Therefore, each correlation of error aspects appears in four separate tables: one for the entire Author/Title catalog, one for the Santa Cruz

records in the Author/Title catalog, one for the entire Subject catalog, and one for the Santa Cruz records in the Subject catalog. Thus, while 48 tables were created, only 12 actual correlations of error aspects are presented.

In considering the reported error rates, we should keep in mind that 152 entries of the sample of 5,900 (2.6%) contained some foreign language words which we could not analyze and which were not thoroughly inspected due to the unavailability of people who could read those languages.

It should also again be mentioned that there are some discrepancies in the tabulated results of the study which may be due to keying errors in the punched cards which were not key verified or to the re-coding of some of the data. Almost all of these discrepancies appear in the Subject catalog data, where the totals of some of the tables vary between 1,194 and 1,171. The few variations in totals in tables for the Author/Title catalog are no greater than 4. Most of the discrepancies are of little statistical significance. The totals in the tables presented here, therefore, reflect those discrepancies; percentages given in the tables are percentages of the total given in that table.

B. DISPLAY AND DISCUSSION OF FINDINGS FOR THE AUTHOR/TITLE CATALOG

1. Summary of Error Rate

The absolute numbers of the errors found in this study are noted in the tables.

	NUMBER OF ERRORS			
	FATAL	SERIOUS	MINOR	TOTAL
Author/Title Catalog	141	1,396	1,630	3,167
	(4.4%)	(44.1%)	(51.5%)	(100.0%)
Subject Catalog	159	490	522	1,171
	(13.6%)	(41.8%)	(44.6%)	(100.0%)
TOTAL	300	1,886	2,152	4,338
	(6.9%)	(43.5%)	(49.6%)	(100.0%)

TABLE 1: TOTAL NUMBER OF ERRORS FOUND IN THE SAMPLE

The estimated catalog error rates can be computed from the above data and the sample size data given earlier. This results in the data shown in Table 2.

ERROR RATE	FATAL	SERIOUS	MINOR	TOTAL
Author/Title Catalog				
Errors Per page	2.3	22.9	26.7	51.9
Errors Per entry	0.04	0.39	0.45	0.88
Subject Catalog				
Errors Per page	4.8	14.8	15.8	35.4
Errors Per entry	0.07	0.21	0.23	0.51
TOTAL				
Errors Per page	3.2	20.0	22.9	46.1
Errors Per entry	0.05	0.32	0.36	0.74

TABLE 2: SUMMARY OF COMPUTED ERROR RATE

2. Causes of Error in the Author/Title Catalog

Figure 1 and Section III C. 4. in this report listed and discussed the causes of error that were considered for this study; all errors were attributed to one of these categories of causes. The gross distribution of total errors (fatal, serious, minor) by cause is given in Table 3.

Table 3 shows the errors in the Author/Title catalog sample aranged according to error type and cause. In each cell of the table we find the number of errors found of a certain type, with a certain cause. For example, we see in the first horizontal row that no errors were found representing "duplicate data not suppressed" which were caused by keeping errors or variant cataloging practice, but 9 were found (not surprisingly) due to program processing failures, and 12 were found due to unknown causes. In all there were 21 errors found in this type category, 43 of which were due to program processing and 57% of which were due to unknown causes.

Some types of errors had consistent causes. For example, all 30 of the transposition errors (Type 31) were due (again, not surprisingly) to keying errors, and all 50 of the instances of non-English articles being used improperly in filing were due to inadequate design. (No attempt was made in the UCUCS programs to disregard non-English articles in filing. Since filing errors resulting from non-English articles could have been suppressed for most languages, these filing errors must be ascribed to inadequate design rather than program processing). We can also see from this table that there were 215 instances of inappropriate entries found in the sample (all due to processing other than programming).

Type of Error	CAUSE OF ERROR: Keying Error(10)	Variant Cataloging Practice(20)	Program Processing(30)	Processing(40)	Inadequate Design(50)	Other(60)	Unknown(70)	Total
Duplicate data not suppressed (10)	0	0	9	0	0	0	12	21(0.7)*
Variant form of same data (20)	0	96	0	0	0	9	3	108(3.4)
Orthographic inaccuracy								
--transposition (31)	30	0	0	0	0	0	0	30(0.9)
--missing string (32)	221	112	410	0	0	1	15	769(24.3)
--added string (33)	91	25	21	0	0	1	10	148(4.7)
--meaningless string (34)	25	0	14	0	0	0	6	45(1.4)
--missing or added blank (35)	555	2	28	0	0	1	11	597(18.8)
--incorrect or missing caps (36)	131	0	11	0	0	43	128	313(9.9)
--other misspelling or undeter. (37)	380	9	32	0	0	0	8	429(13.5)
String improperly used in filing								
--function term (41)	0	8	0	0	100	0	0	108(3.4)
--dates (42)	8	98	0	0	3	0	0	109(3.5)
--associated title (43)	0	0	9	0	21	0	0	30(0.9)
--English article (44)	0	0	0	0	0	0	0	0
--non-English article (45)	0	0	0	0	50	0	0	50(1.6)
--other (46)	0	0	2	0	10	0	2	14(0.5)
Data field missing (50)	0	1	5	0	0	0	48	54(1.7)
Data field added (60)	0	0	101	4	9	0	0	114(3.6)
Inappropriate entry (70)	0	0	0	215	0	0	0	215(6.8)
Other (80)	0	1	2	0	0	1	9	13(0.4)
TOTAL	1,441(45.5)*	362(11.4)	644(20.3)	219(6.9)	193(6.1)	56(1.8)	252(8.0)	3,167(100.0)

*Numbers in parentheses represent percent of total sample.

Table 3

Number and Percent of All Errors in Sample by Type and Cause, Author/Title Catalog

Our study delineated six ways in which character strings could be improperly used in filing: as function terms (e.g., editor, trans.), dates (author or editor dates), associated titles (e.g., Mrs., 1st baron), English articles, and non-English articles, and "Other," encompassing errors not fitting neatly into any of the previous five categories. In all, 311 instances were found of strings improperly used in filing. Function terms and dates were the kind of strings most frequently misused in filing. But there were more than 311 instances of misfiled entries found in the Author/Title portion of the sample. Most of the error type categories as listed in this table can be associated with misfiling of entries, either by causing an entry to appear in the wrong filing position or by causing an entry to fail to consolidate with another entry and thereby also filing in the wrong place.

We can see from this table that keying errors caused the largest percentage of errors (45.5%) found in the Author/Title catalog sample. The next highest percentage (20.3%) was contributed by program processing failures.

It is of interest here to compare the analogous data given in Table 4 for the Santa Cruz records (errors arranged by type and cause).

Here we see that the highest percentage of errors in the Santa Cruz part of the sample was contributed by failures in program processing. It is impossible for us to say which programs were responsible for the errors. Errors could have occurred in the programs which generated the Santa Cruz tapes; they could have occurred in the programs which attempted to carry out any of the operations performed on the entire file (such as sorting, consolidation, generating added entries, etc.). It seems likely, however, that most of the program failure in the Santa Cruz records occurred either in the programs used to produce the tape which was sent to ILR or else in the programs which attempted to merge these records into the rest of the UCUCS file. This is a logical conclusion since 531 of the total 644 program processing errors in the Author/Title catalog occurred in Santa Cruz records. That is, 82.5% of all the program processing errors found in the Author/Title sample occurred in Santa Cruz records.

Comparing Tables 3 and 4 we can also note that of the 252 errors ascribed to "Unknown" causes, 160 or 63.5% occurred in Santa Cruz records. This reflects the subjective observation by the authors that rather bizarre and inexplicable errors occurred more frequently in these records.

Finally, it should be noted that 1,211 of the 3,167 errors found in the Author/Title sample were found in records from Santa Cruz. This represents 38.2% of all the errors found in the Author/Title catalog sample. Analogous tables for the Subject catalog sample are presented in Section C. From these tables we can see that 30.1% of the errors found in the Subject catalog sample were found in Santa Cruz records. For the entire sample of 4,361 errors, 36.0% (1,570) occurred in Santa Cruz records. Santa Cruz contributed a total of 122,240 titles (representing 16.3% of the UCUCS titles, and 11.7% of the UCUCS records), and hence one would expect the errors from UCUCS processing to be distributed over Santa Cruz records at about that same proportion.

	CAUSE OF ERROR							
Type of Error	Keying Error(10)	Variant Cataloging Practice(20)	Program Processing(30)	Processing(40)	Inadequate Design(50)	Other(60)	Unknown(70)	Total
Duplicate data not suppressed (10)	0	0	3	0	0	0	1	4 (0.3)*
Variant form of same data (20)	0	21	0	0	0	9	1	31 (2.5)
Orthographic innacuracy								
--transposition (31)	1	0	0	0	0	0	0	1 (0.1)
--missing string (32)	58	47	399	0	0	1	8	513 (42.3)
--added string (33)	9	2	17	0	0	0	2	30 (2.5)
--meaningless string (34)	3	0	9	0	0	0	0	12 (1.0)
--missing or added blanks (35)	158	1	16	0	0	1	5	181 (14.9)
--incorrect or missing caps (36)	26	0	1	0	0	0	111	138 (11.4)
--other misspelling or undetermined(37)	122	0	12	0	0	0	5	139 (11.5)
String improperly used in filing								
--function term (41)	0	4	0	0	10	0	0	14 (1.2)
--dates (42)	2	13	0	0	0	0	0	15 (1.2)
--associated title (43)	0	0	0	0	2	0	0	2 (0.2)
--English article (44)	0	0	0	0	0	0	0	0
--non-English article (45)	0	0	0	0	3	0	0	3 (0.3)
--other (46)	0	0	0	0	0	0	0	0
Data field missing (50)	0	0	1	0	0	0	27	28 (2.3)
Data field added (60)	0	0	73	3	1	0	0	77 (6.4)
Inappropriate entry (70)	0	0	0	23	0	0	0	23 (1.9)
Other (80)	0	0	0	0	0	0	0	0
TOTAL	379 (31.3)*	88 (7.3)	531 (43.8)	26 (2.1)	16 (1.3)	11 (1.0)	160 (13.2)	1,211 (100.0)

*Numbers in parentheses represent percent of total sample.

Table 4

Number and Percent of All Errors in Sample by Type and Cause, Author/Title Catalog, Santa Cruz Records Only

We can also see that a high proportion of those types of errors called "orthographic inaccuracies" appeared in the Santa Cruz records. For example, there were 769 instances of missing strings found in the entire Author Title sample; 513 (66.7%) of these appeared in Santa Cruz records. There were 597 instances of missing or added blanks in the entire Author/Title sample; 181 (30.3%) of these errors were contributed by Santa Cruz records. The total of all types of orthographic inaccuracies for the Author/Title sample was 2,331 (by far the most frequent general error type--73.6% of all error found in the Author/Title catalog sample). Of these 2,331 errors, 43.5% (1,014) were found in Santa Cruz records. As will be discussed below, most orthographic inaccuracy errors were relatively minor, usually affecting only the appearance of the entry. However, some had more serious effects, such as non-consolidation, misfiling of entries onto other pages, and making the content of the record uncertain.

3. Seriousness of the Errors in the Author/Title Catalog

Let us now consider the question of the severity of the errors found in our sample. We divided the errors into three categories according to the effect each error had. Of course, some errors had more than one determinable effect, but since only one effect could be recorded for each error, the most serious effect was chosen when there was a choice.

Minor errors included only three effect categories: wasted space (the error had no effect more serious than wasting space in the catalog); appearance (the error affected only the appearance of the entry); and other or unknown. The other or unknown category was included in the minor errors because we believed that the categories for the more serious kinds of errors had been carefully enough defined so that little, if anything, had been left out. Table 5 displays the minor errors according to their causes and effects. Here we see that there were 1,630 minor errors in the Author/Title catalog sample, or 51.5% of all the errors in the Author/Title segment of the sample. Most of these errors (902 or 55.3%) were caused by keying errors, 18.7% were caused by program processing, and 13.4% were caused by record processing other than keying or programming.

From Table 6, displaying the data for the minor errors in the Santa Cruz records in the Author/Title segment, we can see that 665 or 40.8% of the 1,630 minor errors found in the entire Author/Title segment were found in Santa Cruz records. For the Santa Cruz portion of this segment, program processing caused nearly as many minor errors as did keying mistakes: 274 (41.2%) were caused by keying errors and 246 (37.0%) were caused by program processing.

The serious errors included just two general types of error: those which resulted in uncertainty of the content of the record, and those which resulted in the misfile of the record in the same column or same page (that is, the record was misfiled but still appeared on the same page where it was supposed to appear). Both of these types of errors can, in some cases, prevent a user from finding a needed item. But these errors really represent differing degrees of seriousness. Misfiling of an entry onto another column of the same page is more likely to result in the user missing that entry point than is misfiling within the same column, particularly if the needed item appears immediately adjacent to the entry

CAUSE OF MINOR ERRORS

Effect of Minor Errors	Keying Error(10)	Variant Cataloging Practice(20)	Program Processing(30)	Processing(40)	Inadequate Design(50)	Other(60)	Unknown(70)	Total
Wasted space (40)	31	0	84	219	8	0	10	352(21.6)*
Appearance only (80)	869	1	183	0	0	44	135	1,232(75.6)
Other or unknown (90)	2	1	38	0	1	0	4	46(2.8)
TOTAL	902(55.3)*	2(0.1)	305(18.7)	219(13.4)	9(0.6)	44(2.7)	149(9.1)	1,630(100.0)

*Numbers in parentheses represent percent of total sample.

Table 5

Number and Percent of All Minor Errors in Sample by Cause and Effect, Author/Title Catalog

CAUSE OF MINOR ERRORS

Effect of Minor Errors	Keying Error(10)	Variant Cataloging Practice(20)	Program Processing(30)	Processing(40)	Inadequate Design(50)	Other(60)	Unknown(70)	Total
Wasted space (40)	2	0	66	26	1	0	0	95(14.3)*
Appearance only (80)	272	0	165	0	0	1	116	554(83.3)
Other or unknown (90)	0	0	15	0	0	0	1	16(2.4)
TOTAL	274(41.2)*	0	246(37.0)	26(3.9)	1(0.2)	1(0.2)	117(17.6)	665(100.0)

*Numbers in parentheses represent percent of total sample.

Table 6

Number and Percent of All Minor Errors in Sample by Cause and Effect, Author/Title Catalog, Santa Cruz Records Only

point where it should be. And since we interpreted the category "content uncertain" so strictly that errors would be so categorized if they presented any doubt about the meaning of the word or data element, many errors tallied under this category might better have been included with the minor errors.

Placing these errors in the serious category was an arbitrary choice in keeping with the operating rule of "When in doubt, don't give the benefit of the doubt to the catalog." The summary of all serious errors found in the Author/Title sample is presented in Table 7. As noted in Table 1 serious errors represent 44.1% of the errors in the Author/ Title segment of the sample. Most of the serious errors were caused by keying errors (509 of them, or 36.5%). Variant cataloging practice and program processing nearly tie for next most frequent cause (23.4% and 23.9%, respectively). We notice that almost half of all the serious errors recorded had the effect of content uncertain (656 of 1,396, or 47.0%). Given the arbitrary placement of this category into the serious error group, it is useful to consider what the results would be if the errors tallied under "content uncertain" were omitted from the table. Without the content uncertain errors, there would be a total of 740 serious errors in the Author/Title segment, or 23.4% of the 3,167 errors in this segment of the sample. Also, deleting these errors from the tabulated data changes the relative frequency of causes of serious errors. Of the 509 serious errors caused by keying errors, 312 would be deleted, leaving 197 serious errors caused by keying mistakes. This figure represents 26.6% of all these errors. The variant cataloging practice errors would have nearly the same total as in Table 7 but a higher relative frequency: 319 instead of 327, 43.1% instead of 23.4%. Program processing would then cause only 71 of the serious errors and would be the cause of serious errors only 10.0% of the time instead of the 23.9% given in the table.

In summary, then, if those errors having the effect of content uncertain were considered minor instead of serious, the most prevalent cause of serious errors would be variant cataloging practice rather than keying errors, and program processing would have caused only 10% rather than nearly 24% of the serious errors. We won't list details of the effects of adding these content uncertain errors to the minor error tables. The interested reader can do that easily enough. It does seem worth mentioning, however, that including the content uncertain errors with the other minor errors would increase the relative frequency of minor errors in the Author/Title segment of the sample from 51.5% to 72.2%. The serious and fatal errors together would then equal 27.9% rather than 48.6% of the Author/Title segment of the sample.

The Santa Cruz portion of the sample accounts for 522 or 37.4% of the total 1,396 serious errors in the Author/Title catalog, as is shown in Table 8. The sample also accounts for over half of the 656 serious errors in the Author/Title catalog which result in uncertain content, and for about 85% of the serious errors in the Author/Title catalog which are caused by program processing. Over half (54.2%) of the serious errors in the Santa Cruz sample were accounted for by program processing, with keying

CAUSE OF SERIOUS ERRORS

Effect of Serious Errors	Keying Error(10)	Variant Cataloging Practice(20)	Program Processing(30)	Processing(40)	Inadequate Design(50)	Other(60)	Unknown(70)	Total
Misfiled in same column w/heading (21)	5	17	3	0	3	0	2	30(2.1)*
Misfiled on same page w/heading (22)	2	2	8	0	7	0	0	19(1.4)
Misfiled in same column without heading (24)	0	0	0	0	3	0	0	3(0.2)
Misfiled on same page without heading (25)	0	0	0	0	0	0	0	0
Content uncertain (30)	312	8	262	0	0	2	72	656(47.0)
Non-consolidated entry								
--in same column w/heading (51)	66	71	9	0	3	0	11	160(11.5)
--on same page w/heading (52)	9	7	1	0	1	0	0	18(1.3)
--in same column without heading (54)	93	107	47	0	0	0	14	261(18.7)
--on same page without heading (55)	3	12	3	0	2	0	0	20(1.4)
Non-consolidated heading								
--in same column w/heading (61)	16	46	0	0	53	0	0	115(8.2)
--on same page with heading (62)	2	6	0	0	8	0	0	16(1.2)
Non-consolidated heading								
--in same column without heading (64)	1	30	0	0	33	0	0	64(4.6)
--on same page without heading (65)	0	21	0	0	13	0	0	34(2.4)
TOTAL	509(36.5)*	327(23.4)	333(23.9)	0	126(9.0)	2(0.1)	99(7.1)	1,396(100.0)

*Numbers in parentheses represent percent of total sample.

Table 7

Number and Percent of All Serious Errors in Sample by Cause and Effect, Author/Title Catalog

	CAUSE OF SERIOUS ERROR						
Effect of Serious Errors	Keying Error(10)	Variant Cataloging Practice(20)	Program Processing(30)	Inadequate Design(50)	Other(60)	Unknown(70)	Total
Content uncertain (30)	62	1	230	0	1	36	330(63.2)*
Non-consolidated entry							
--in same column w/heading (51)	16	27	8	0	0	1	52(10.0)
--on same page w/heading (52)	0	2	1	0	0	0	3(0.6)
--in same column without heading (54)	20	41	43	0	0	6	110(21.1)
--on same page without heading (55)	1	2	1	0	0	0	4(0.8)
Non-consolidated heading							
--in same column w/heading (61)	2	4	0	6	0	0	12(2.3)
--on same page with heading (62)	0	1	0	0	0	0	1(0.2)
Non-consolidated heading							
--in same column without heading (64)	0	2	0	3	0	0	5(0.9)
--on same page without heading (65)	0	3	0	2	0	0	5(0.9)
TOTAL	101(19.4)*	83(15.9)	283(54.2)	11(2.1)	1(0.2)	43(8.2)	522(100.0)

*Numbers in parentheses represent percent of total sample.

Table 8

Number and Percent of All Serious Errors in Sample by Cause and Effect, Author/Title Catalog, Santa Cruz Records Only

errors (19.4%) and variant cataloging practice (15.9%) being the next most frequent causes. We can see also that 330 or 63.2% of the serious errors in the Santa Cruz sample had the effect of contents uncertain and of these 330 errors, 230 (69.7%) were caused by program processing problems. Deleting the contents uncertain category in this case would reduce the total number of serious errors in the Author/Title portion of the Santa Cruz sample to 192. Program processing errors would then account for 53 or only 28.1% of this total, and keying errors accounting for 39 or 20.3%. Variant cataloging practice would then be the cause of the largest number (82, or 42.9%) of serious errors in the Santa Cruz sample.

The fatal error category was, fortunately, less problematic than the serious error category. The serious errors consist of those errors which definitely result in a lost entry point (effect 10--lost entry point) and all those which are very likely to result in a lost entry point (all those involving a misfile onto another page). Table 9 provides data on the cause and effect of fatal errors in the Author/Title catalog. One of the most interesting features to be noticed in this table is that, for the first time, inadequate design is responsible for the plurality of errors. Here, inadequate design has contributed 58 of the 141 fatal errors, or 41.1%. Variant cataloging practice is second, with 23.4%, and keying errors run a close third (21.3%). Remembering that keying errors contributed the overwhelming plurality of all errors found in the Author/Title catalog (45.5% versus 20.3% for the second most frequent cause), it is interesting to note here that it is a less significant factor in the fatal error causes than is inadequate design and that it is approximately equal in frequency with variant cataloging practice.

It is also worthwhile to point out in Table 10 that in the case of fatal errors Santa Cruz records do not contribute a significantly high percentage of the errors. Of the 141 fatal errors in the Author/Title sample, 24 (or 17.0%) were in Santa Cruz records. The numbers in this table are so small that little else can be concluded from them. However, it does seem worth noticing that close to half (41.7%) of these errors were lost entry points due to causes other than the five specifically defined cause categories.

CAUSE OF FATAL ERROR

Effect of Fatal Error	Keying Error(10)	Variant Cataloging Practice(20)	Program Processing(30)	Processing(40)	Inadequate Design(50)	Other(60)	Unknown(70)	Total
Lost Entry point (10)	3	0	4	0	0	10	3	20(14.2)*
Misfiled off the page with heading (23)	7	0	0	0	49	0	0	56(39.7)
Misfiled off the page without heading (26)	0	0	0	0	0	0	0	0
Non-consolidation of entry off the page with heading (53)	3	4	1	0	0	0	0	10(7.1)
Non-consolidation of entry off the page without heading (56)	2	6	1	0	0	0	0	9(6.4)
Non-consolidation of heading off the page with heading (63)	6	4	0	0	5	0	1	16(11.3)
Non-consolidation of heading off the page without heading (66)	7	19	0	0	4	0	0	30(21.3)
TOTAL	30(21.3)*	33(23.4)	6(4.3)	0	58(41.1)	10(7.1)	4(2.8)	141(100.0)

*Numbers in parentheses represent percent of total sample.

Table 9

Number and Percent of All Fatal Errors in Sample by Cause and Effect, Author/Title Catalog

CAUSE OF FATAL ERROR

Effect of Fatal Error	Keying Error(10)	Variant Cataloging Practice(20)	Program Processing(30)	Processing(40)	Inadequate Design(50)	Other(60)	Unknown(70)	Total
Lost entry point (10)	1	0	0	0	0	9	0	10(41.7)*
Misfiled off the page with heading (23)	0	0	0	0	3	0	0	3(12.5)
Misfiled off the page without heading (26)	0	0	0	0	0	0	0	0
Non-consolidation of entry off the page with heading (53)	1	0	1	0	0	0	0	2(8.3)
Non-consolidation of entry off the page without heading (56)	0	1	1	0	0	0	0	2(8.3)
Non-consolidation of heading off the page with heading (63)	1	0	0	0	1	0	0	2(8.3)
Non-consolidation of heading off the page without heading (66)	1	4	0	0	0	0	0	5(20.9)
TOTAL	4(16.7)*	5(20.8)	2(8.3)	0	4(16.7)	9(37.5)	0	24(100.0)

*Numbers in parentheses represent percent of total sample.

Table 10

Number and Percent of all Fatal Errors in Sample by Cause and Effect, Author/Title Catalog, Santa Cruz Records Only

4. Location of Appearance of the Errors in the Author/Title Catalog

We have so far discussed the error aspects of cause, type, and effect. Another question which might be asked--particularly by those people considering file improvement--is, "Where in the records are the errors located?" This is an important question for people interested in correcting errors in the file, since different error correction devices can be used on various parts of the records. For example, an authority file for author names can be used for finding and correcting errors in author names, an English language dictionary authority list can be used on English language titles, a subject heading authority list can be used on subject headings, and so on. It was hoped that this study would assist those involved in file improvement to decide which error correction devices might best be employed.

As mentioned in the methodology section of this report, there was sometimes a problem when errors were found in added entries in the Author/Title catalog or in entries in the Subject catalog: we could see that an error had occurred (for example, failure of entries to consolidate correctly), but the cause and source location of the error could not be determined without looking at the full bibliographic record in the Author/Title catalog. We wanted to tally errors according to where they appeared for various reasons, but primarily in order to estimate accurately the amount of space wasted in the catalog due to inappropriate entries and non-consolidation of entries. But people interested in file improvement would probably be more concerned with the source location of the error (that is, where in the full bibliographic record the error appears) and not so interested in where the error appears in the subject or added entry. Therefore, we recorded the data so that both kinds of error location could be tallied in data reduction.

Let us consider, in Table 11, the Author/Title sample according to cause of error and location of error appearance. First we note that of the total errors found in the Author/Title catalog sample, over twice as many were found in main entries (2,161) as were found in added entries (1,002). This makes sense because there are more data elements appearing in the main entries, and therefore more opportunities for errors to appear.

We can see from this table that more errors (532) occurred in the collation statements of main entries than in any other portion of the entries in the Author/Title segment of the sample. (The influence of the Santa Cruz records on this figure will be discussed below.) The next most frequent location of error appearance was in added entry headings (421). Most added entry headings are generated from the title statement of main entries (including short title, author statement, editor statement, etc.). It is therefore not surprising that the next most frequent location for errors to appear was in those elements which make up the title statment of the main entries. The total number of errors found in the title statement of main entries was 368.

The reader may notice that a total of 147 errors were found in the call numbers and/or location codes of the entries in this part of the sample. This might seem alarming if we assume that errors in call numbers or location codes would necessarily lead the catalog user astray. Fortunately, this is not the case. Most errors found in these parts of the entries were very minor--such as missing or added blanks--and generally affected only the appearance of the record.

Location of Appearance of Error	CAUSE OF ERROR: Keying Error(10)	Variant Cataloging Practice(20)	Program Processing(30)	Processing(40)	Inadequate Design(50)	Other(60)	Unknown(70)	Total
Main entry (10)	2	1	3	134	0	0	9	149(4.7)*
Main entry--entry heading (11)	83	123	10	0	62	3	4	285(9.0)
--title statement (12)	4	1	0	0	2	0	2	9(0.3)
--title statement, short title (41)	128	18	4	0	1	0	8	159(5.0)
--title statement, subtitle (42)	68	5	2	0	0	0	1	76(2.4)
--title statement, author statement(43)	32	40	2	0	0	0	1	75(2.4)
--title statement, other (44)	43	2	3	0	0	0	1	49(1.6)
--edition statement (13)	11	4	14	0	1	0	1	31(1.0)
--place (14)	24	1	11	0	0	0	9	45(1.4)
--publisher (15)	57	7	12	0	0	0	13	229(7.2)
--date (16)	13	7	60	0	0	0	11	91(2.9)
--collation (17)	312	2	198	0	0	2	18	532(16.8)
--notes (18)	113	0	2	0	0	0	2	117(3.7)
--tracings (19)	142	0	2	0	0	1	121	266(8.4)
--call number/location (45)	16	0	14	0	0	0	6	36(1.1)
Main entry--other (46)	9	0	0	0	0	0	3	12(0.4)
Added entry (20)	24	37	11	80	0	0	13	165(5.2)
Added entry--headings (21)	171	68	13	1	111	50	7	421(13.3)
--title statement (22)	51	0	10	0	7	0	3	71(2.2)
--date (23)	2	1	20	0	0	0	7	30(1.0)
--see reference (24)	48	44	13	0	0	0	5	110(3.5)
--call number/loc. code (25)	81	1	5	0	0	0	7	94(3.0)
--other (26)	6	0	93	3	9	0	0	111(3.5)
TOTAL	1,440(45.5)*	362(11.4)	642(20.3)	218(6.9)	193(6.1)	56(1.8)	252(8.0)	3,163(100.0)

*Numbers in parentheses represent percent of total sample

Table 11

Number and Percent of All Errors in Sample by Location of Error Appearance and Cause, Author/Title Catalog

It is worthwhile to discuss briefly the corresponding data in Table 12 for the Santa Cruz subset of this segment of the sample. One of the most interesting points illustrated by comparing this table with the preceding one is that 311 of the 532 errors appearing in collation statements were in Santa Cruz records. This represents 58.5% of all the collation statement errors in the Author/Title sample. Such errors may be a significant problem, because many errors in collation statements were responsible for nonconsolidation (and therefore for misfiling) of entries.

Of the errors found in the whole Author/Title sample in publisher statement and tracings segments of main entries, similarly high percentages were contributed by Santa Cruz records. In publisher statements, 175 of the 228, or 76.8%, of the errors found were in Santa Cruz records. For errors found in tracings, the figures are 168 out of 266, or 63.2%. Although there were relatively few (91) errors found in publication dates, 79.1% (72) of these occurred in Santa Cruz records. We might also note that 77 of the 111 (or 69.4%) of the errors found in added entry elements other than those specifically defined came from Santa Cruz records. This figure reflects the relatively high number of instances of unnecessarily (and incorrectly) added data fields found in Santa Cruz records.

Table 13 correlates errors in the Author/Title Catalog by type and location of error appearance. We see that the most common error type was that of missing string (769, or 24.3% of all errors in the Author/Title segment). The next most frequent error type was that of missing or added blanks (597, or 18.9%). Third and fourth most frequent error types were other or undetermined misspelling and incorrect or missing capitalization, respectively. The four most common error types were all instances of orthographic inaccuracy, which might be expected since we have already learned that keying errors were the most common cause of error (45.5%) and were generally the cause of orthographic inaccuracies. All types of orthographic inaccuracy combined contributed 2,326, or 54.9% of all the errors found in the Author/Title segment.

5. Location of Origin of Errors in the Author/Title Catalog

Now let us consider, in Table 14, the errors according to where they originated in the source records. The figures in Table 14 table differ from Table 11 because 75 of the errors tallied in the general category for added entries in Table 11 have been subtracted from that category and tallied in various categories for main entry locations. This was done because the cause of error in 75 of the added entries was not determinable without examining the corresponding main entries. This means that the source location of these errors was not apparent from the added entries themselves. By comparing this table with Table 11 we see that no dramatic differences appear; these errors were rather insignificantly dispersed through various source locations. One interesting comparison we can note, however, is that in considering location of error appearance, 40 of the 361 (or 10.1%) errors caused by variant cataloging practice occurred in the author part of title statements in main entries. When source location is considered instead, 55 (or 15.2%) of these 362 errors occurred in the author part of main entry title statements.

Location of Appearance of Error	CAUSE OF ERROR: Keying Error(10)	Variant Cataloging Practice(20)	Program Processing(30)	Processing(40)	Inadequate Design(50)	Other(60)	Unknown(70)	Total
Main entry (10)	0	0	0	15	0	0	1	16(1.3)*
Main entry--entry heading (11)	15	19	1	0	10	0	0	45(3.7)
--title statement (12)	0	1	0	0	0	0	0	1(0.1)
--title statement, short title (41)	19	2	4	0	0	0	2	27(2.2)
--title statement, subtitle (42)	25	0	0	0	0	0	1	26(2.1)
--title statement, author statement (43)	9	22	2	0	0	0	0	33(2.7)
--title statement, other (44)	13	0	0	0	0	0	0	13(1.1)
--edition statement (13)	4	3	13	0	0	0	1	21(1.8)
--place (14)	0	0	11	0	0	0	5	16(1.3)
--publisher (15)	19	2	149	0	0	0	6	176(14.6)
--date (16)	4	2	58	0	0	0	8	72(6.0)
--collation (17)	103	0	195	0	0	2	11	311(25.8)
--notes (18)	47	0	1	0	0	0	2	50(4.1)
Main entry--tracings (19)	54	0	2	0	0	0	112	168(13.9)
--call number/location code (45)	2	0	9	0	0	0	4	15(1.2)
--other (46)	3	0	0	0	0	0	0	3(0.2)
Added entry (20)	7	16	11	8	0	0	1	43(3.6)
Added entry--headings (21)	25	7	2	0	5	9	1	49(4.1)
--title statement (22)	17	0	0	0	0	0	0	17(1.4)
--date (23)	0	0	0	0	0	0	3	3(0.3)
--see reference (24)	3	14	0	0	0	0	0	17(1.4)
--call number/location code (25)	4	0	2	0	0	0	2	8(0.7)
--other (26)	5	0	69	2	1	0	0	77(6.4)
TOTAL	378(31.4)*	88(7.2)	529(43.8)	25(2.1)	16(1.3)	11(0.9)	160(13.3)	1,207(100.0)

*Numbers in parentheses represent percent of total sample.

Table 12

Number and Percent of all Errors in Sample by location of Error Appearance and Cause, Author/Title Catalog, Santa Cruz Records Only

Location of Appearance of Error	ORTHOGRAPHIC INACCURACY: Duplicate Data not Suppressed (10)	Variant Form of Name Data(20)	Transposition(31)	Missing String (32)	Added String (33)	Meaningless String (34)	Missing or Added Blanks (35)	Incorrect or Missing Code(36)	Other Misspellings or Undetermined (37)	ERRING IMPROPERLY USED IN FILING: Function Term(41)	Dates (42)	Associated Title (43)	Non-English Article (45)	Other (46)	Data Field Missing (50)	Data Field Added (60)	Inappropriate Entry (70)	Other (80)	Total
Main entry (10)	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	134	4	149(4.7)*
Main entry--entry heading (11)	0	20	2	46	18	1	11	7	28	40	81	28	1	1	0	0	0	1	285(9.0)
--title statement (12)	0	1	1	0	0	0	1	0	2	0	0	0	2	0	1	0	0	1	9(0.3)
--title statement, short title (41)	0	6	4	37	21	3	27	7	49	0	0	0	1	1	0	0	0	3	159(5.0)
--title statement, subtitle (42)	0	3	3	13	5	2	26	3	21	0	0	0	0	0	0	0	0	0	76(2.4)
--title statement, author statement (43)	0	0	1	33	6	1	15	5	6	0	0	0	0	0	0	0	0	0	75(2.4)
--title statement, other (44)	1	1	2	9	3	1	15	3	14	0	0	0	0	0	0	0	0	0	49(1.6)
--edition statement (13)	0	1	0	20	1	0	4	1	2	0	0	0	0	1	1	0	0	0	31(1.0)
--place (14)	0	0	1	13	3	2	6	6	6	0	0	0	0	0	0	0	0	0	45(1.4)
--publisher (15)	0	0	1	164	8	3	4	18	19	0	0	0	0	0	12	0	0	0	229(7.2)
--date (16)	0	1	0	59	4	7	7	0	3	0	0	0	0	0	10	0	0	0	91(2.9)
--collation (17)	0	0	3	193	22	2	257	9	38	0	0	0	0	0	0	0	0	0	532(16.8)
--notes (18)	0	0	2	11	3	1	44	12	43	0	0	0	0	0	1	0	0	0	117(3.7)
--tracings (19)	0	0	4	19	15	7	4	159	56	0	0	0	0	0	1	0	0	1	266(8.4)
--call number/loc. code (45)	3	0	0	0	1	1	12	9	10	0	0	0	0	0	0	0	0	0	36(1.1)
--other (46)	0	0	0	1	1	3	2	1	4	0	0	0	0	0	0	0	0	0	12(0.4)
Added entry (20)	7	5	0	44	7	0	16	0	5	0	0	0	0	0	0	1	80	0	165(5.2)
Added entry--heading (21)	0	25	2	62	18	2	47	56	61	61	26	2	45	11	1	1	1	0	421(13.3)
--title statement (22)	1	1	2	20	4	4	9	3	15	6	0	0	1	0	3	1	0	1	71(2.2)
--date (23)	0	0	1	4	1	2	1	0	15	0	0	0	0	0	6	0	0	0	30(1.0)
--see reference (24)	0	34	1	17	4	2	14	8	21	1	2	0	0	0	1	3	0	2	110(3.5)
--call number/loc. code (25)	1	1	0	4	3	0	73	4	7	0	0	0	0	0	1	0	0	0	94(3.0)
--other (26)	0	0	0	0	0	1	1	1	3	0	0	0	0	0	0	105	0	0	111(3.5)
TOTAL	21(0.7)*	108(3.4)	30(1.0)	769(24.3)	148(4.7)	45(1.4)	597(18.9)	313(9.9)	428(13.5)	108(3.4)	109(3.4)	30(1.0)	50(1.6)	14(0.4)	54(1.7)	111(3.5)	215(6.8)	13(0.4)	3,163(100.0)

*Numbers in parentheses represent percent of total sample

TABLE 13

Number and Percent of All Errors in Sample by Type and Location of Error Appearance, Author/Title Catalog

	CAUSE OF ERROR							
Location of Source of Error	Keying Error(10)	Variant Cataloging Practice(20)	Program Processing(30)	Processing(40)	Inadequate Design(50)	Other(60)	Unknown(70)	Total
Main entry (10)	2	1	3	134	0	0	9	149(4.7)*
Main entry--entry heading (11)	84	123	10	0	62	3	4	286(9.1)
--title statement (12)	4	1	0	0	2	0	2	9(0.3)
--title statement, short title (41)	135	19	4	0	1	0	8	167(5.3)
--title statement, subtitle (42)	73	5	3	0	0	0	1	82(2.6)
--title statement, author statement (43)	33	55	3	0	0	0	5	96(3.0)
--title statement, other (44)	43	4	4	0	0	0	1	52(1.6)
--edition statement (13)	11	12	14	0	1	0	1	39(1.2)
--place (14)	26	7	11	0	0	0	10	54(1.7)
--publisher (15)	60	10	154	0	0	0	13	237(7.5)
--date (16)	13	7	65	0	0	0	11	96(3.0)
--collation (17)	316	3	198	0	0	2	18	537(17.0)
--notes (18)	114	0	2	0	0	0	2	118(3.7)
Main entry--tracings (19)	142	0	2	0	0	1	121	266(8.4)
--call number/location code (45)	16	0	14	0	0	0	6	36(1.1)
--Other (46)	9	0	0	0	0	0	3	12(0.4)
Added entry (20)	0	1	1	80	0	0	8	90(2.9)
Added entry--headings (21)	171	68	13	1	111	50	7	421(13.3)
--title statement (22)	51	0	10	0	7	0	3	71(2.2)
--date (23)	2	1	20	0	0	0	7	30(1.0)
--see reference (24)	48	44	13	0	0	0	5	110(3.5)
--call number/location code (25)	81	1	5	0	0	0	7	94(3.0)
--other (26)	6	0	93	3	9	0	0	111(3.5)
TOTAL	1,440(45.5)*	362(11.4)	642(20.3)	218(6.9)	193(6.1)	56(1.8)	252(8.0)	3,163(100.0)

*Numbers in parentheses represent percent of total sample

Table 14

Number and Percent of All Errors in Sample by Location of Error Source and Cause, Author/Title Catalog

The source location data for the Santa Cruz sample is presented in Table 15. As is true for the subject sample as a whole, few significant trends can be found in the data. It may be noted, however, that 29 (32.9%) of the 88 errors due to variant cataloging practice originated in the author part of the title statement, although only 22 (25%) appeared there.

Analogous to the immediately preceding tables, tables were also generated correlating type with source location of errors. No dramatic differences were found in the two ways of tabulating error location.

6. Other Correlations of Errors in the Author/Title Catalog

So far we have correlated error cause with the other three main error aspects (type, location, and effect); we have correlated error type with two other aspects (cause and location); and we have correlated error location with two other aspects (cause and type). Error effect has only been correlated with cause in the tables describing minor, serious and fatal errors. The effect aspect has not been correlated with error location or error type, but the source data is available to permit this to be done at a later date if desired.

7. Effect of Non-Consolidation of Entries in the Author/Title Catalog

Non-consolidation occurred in various forms in the catalog. It could occur in entry headings only, in the bodies of entries only, or in both; these three categories of non-consolidation could occur in main entries, added entries, and subject entries. Subject entries could also have non-consolidated subject headings, with or without any of the other kinds of non-consolidation. There are a total of 17 possible combinations of these factors.

The various types of non-consolidation involved differing degrees of space wasted in the catalog. For example, the non-consolidation of two main entries might result in two inches of column space being wasted, whereas the non-consolidation of two added entries might waste only 1/4-inch of space, and the non-consolidation of an entry heading only might waste even less space in the catalog. Consequently, in order to arrive at an accurate estimate of the amount of space wasted in the catalog, it would be necessary to consider all 17 of the types of non-consolidation and estimate the amount of space wasted by each one.

The effect category, wasted space, was used for all instances of inappropriate entries and also for instances of added data elements which used up at least one extra line of type. Inappropriate entries could occur as main, added, and subject entries. The type of entry would affect the amount of space wasted. Added data elements could appear in either added or subject entries. It is assumed that the type of entry in which an added data element appears would not substantially affect the amount of space wasted.

Detailed study of wasted space in UCUCS would involve a complex analysis of all these elements, an analysis prevented by the time limitations on this project. A very rough estimate of space wasted in UCUCS due to non-consolidation of entries was made in an unpublished student paper by Judy Todd and others for a systems analysis class in the School of Librarianship, U.C. Berkeley. For that paper a sample of 15 pairs of pages was xeroxed from Volume I (A-Ana) of the Author/Title catalog of UCUCS. These 30 pages

Location of Source of Error	CAUSE OF ERROR: Keying Error(10)	Variant Cataloging Practice(20)	Program Processing(30)	Processing(40)	Inadequate Design(50)	Other(60)	Unknown(70)	Total
Main entry (10)	0	0	0	15	0	0	1	16(1.3)*
Main entry--entry heading (11)	15	19	1	0	10	0	1	46(3.8)
--title statement (12)	0	1	0	0	0	0	0	1(0.1)
--title statement, short title (41)	22	2	4	0	0	0	2	30(2.5)
--title statement, subtitle (42)	27	0	1	0	0	0	1	29(2.4)
--title statement, author statement (43)	9	29	3	0	0	0	0	41(3.4)
--title statement, other (44)	13	1	1	0	0	0	0	15(1.2)
--edition statement (13)	4	6	13	0	0	0	1	24(2.0)
--place (14)	1	3	11	0	0	0	5	20(1.7)
--publisher (15)	20	4	151	0	0	0	6	181(15.0)
--date (16)	4	2	63	0	0	0	8	77(6.4)
--collation (17)	103	0	195	0	0	2	11	311(25.8)
--notes (18)	47	0	1	0	0	0	2	50(4.1)
Main entry--tracings (19)	54	0	2	0	0	0	112	168(13.9)
--call number/location code (45)	2	0	9	0	0	0	4	15(1.2)
--Other (46)	3	0	0	0	0	0	0	3(0.2)
Added entry (20)	0	0	1	8	0	0	0	9(0.8)
Added entry--headings (21)	25	7	2	0	5	0	1	49(4.1)
--title statement (22)	17	0	0	0	0	0	0	17(1.4)
--date (23)	0	0	0	0	0	0	3	3(0.2)
--see reference (24)	3	14	0	0	0	0	0	17(1.4)
--call number/location code (25)	4	0	2	0	0	0	2	8(0.7)
--other (26)	5	0	69	2	1	0	0	77(6.4)
TOTAL	378(31.3)*	88(7.3)	529(43.8)	25(2.1)	16(1.3)	11(0.9)	160(13.3)	1,207(100.0)

*Numbers in parentheses represent percent of total sample.

Table 15

Number and Percent of All Errors in Sample by Location of Error Source and Cause, Author/Title Catalog, Santa Cruz records only

were searched for main and added entry non-consolidation of entry heading only and of the entire entry. Space wasted due to other added data elements was not considered in this report.

Of the sample of 1,895 entries, 16.1% were duplicate entries, that is, redundant entries which should have consolidated with other entry headings or whole entries. These duplicates were analyzed according to the following categories: type of entry (added entry or main entry), area of duplication discrepancies, typographical errors, lack of Automatic Format Recognition program, undetermined). Results of this particular study indicated an average of about 3 column-inches per page or about 10.4% of the printed column-space were taken up by duplicate entries. As the sample size for this study was small and as added data elements were not considered in rough and possibly conservative estimate of wasted space in the UCUCS catalog because of failure to consolidate entries.

C. DISPLAY AND DISCUSSION OF FINDINGS FOR THE SUBJECT CATALOG

1. Summary of Error Rate

The summary data for the number of errors found in the Subject catalog during this study are repeated here from Table 1.

	NUMBER OF ERRORS			
	FATAL	SERIOUS	MINOR	TOTAL
Subject Catalog	159(13.6%)	490(41.8%)	522(44.6%)	1,171(100.0%)

The estimated Subject Catalog error rates are repeated here from Table 2.

	ERROR RATE			
	FATAL	SERIOUS	MINOR	TOTAL
Errors per page	4.8	14.8	15.8	35.4
Errors per entry	0.07	0.21	0.23	0.51

It may be repeated here that there were discrepancies in totals of some of the tables, particularly for the Subject section of the sample. For instance, tables correlating cause and effect for minor, serious, and fatal errors total 1,171, but the table correlating cause and type for the Subject sample totals 1,193. The discrepancies have no great statistical significance. Percentages within tables presented here are percentages of the total given in that particular table.

2. Causes of Errors in the Subject Catalog

Figure 1 and Section III. C. 4. listed and discussed the causes that were to be considered for this study, and all errors were attributed to one of these categories of causes. The gross distribution of total errors (fatal, serious, and minor) by cause is given in Table 16.

We can see from this table that keying errors caused the largest percentage of errors (39.0%) found in the Subject catalog sample. The next highest percentage (22.4%) was contributed by program processing failure. These findings are consistent with those for the Author/Title catalog, where keying errors also contributed the highest percentage (45.5%) and program processing failure the next highest percentage (20.3%) of errors.

The most common error type in the Subject catalog was found to be orthographic inaccuracy, which totalled 718, or 60.2%, of the 1,193 errors in the Subject sample. (Orthographic inaccuracy was also the most common error type in the Author/Title catalog, where 73.6% of the errors were of

Type of Error	CAUSE OF ERROR: Keying Error (10)	Variant Cataloging Practice (20)	Program Processing (30)	Processing (40)	Inadequate Design (50)	Unknown (70)	Total
Duplicate data not suppressed (10)	0	0	9	0	0	15	24(2.0)*
Variant form of same data (20)	0	35	0	0	0	0	35(2.9)
Orthographic inaccuracy							
--transposition (31)	27	18	0	0	1	1	47(3.9)
--missing string (32)	106	43	33	4	2	13	201(16.8)
--added string (33)	44	8	5	0	0	74	131(11.0)
--meaningless string (34)	11	0	3	0	1	4	19(1.6)
--missing or added blank (35)	67	2	1	0	0	0	70(5.9)
--incorrect or missing caps (36)	15	0	2	0	0	8	25(2.1)
--other misspelling or undeter. (37)	193	5	21	0	1	5	225(18.9)
String improperly used in filing							
--function term (41)	0	1	0	0	4	0	5(0.4)
--dates (42)	1	16	1	0	1	0	19(1.6)
--associated title (43)	0	0	0	0	4	0	4((0.3)
--English article (44)	-	-	-	-	-	-	-
--non-English article (45)	0	0	0	0	3	0	3(0.3)
--other (46)	0	0	1	0	0	0	1(0.1)
Data field missing (50)	0	1	21	0	0	5	27(2.3)
Data field added (60)	2	1	168	3	41	2	217(18.2)
Inappropriate entry (70)	0	0	1	92	0	0	93(7.8)
Other (80)	0	0	1	0	0	46	47(3.9)
TOTAL	466(39.0)*	130(10.9)	267(22.4)	99(8.3)	58(4.9)	173(14.5)	1,193(100.0)

*Numbers in parentheses represent percent of total sample.

Table 16

Number and Percent of All Errors in Sample by Type and Cause, Subject Catalog

this error type.) As can be seen in Table 17, (errors in Santa Cruz records by type and cause), keying errors were most commonly responsible for orthographic inaccuracies, accounting for 64.5% of this error type in the Subject sample. As was mentioned in the discussion of such errors in the Author/Title catalog (Section IV. B. 2.), orthographic inaccuracies usually have only minor effects, but occasionally may be responsible for more serious errors, such as non-consolidation of entries.

The Santa Cruz sample had characteristics somewhat different from the sample as a whole. Here, for instance, most errors (46.8%) were due to program processing failure rather than keying mistakes; keying errors do account for the next highest percentage (26.5%) of errors, however. As for the Subject catalog as a whole, the most common error type was orthographic inaccuracy, which accounts for 162, or 45.1%, of the 359 errors in the Santa Cruz Subject catalog sample, and over half (56.8%) of these orthographic inaccuracies are the result of keying errors. However, the Santa Cruz subset, unlike the Subject Catalog as a whole, has a significant percentage (39.8%) of errors categorized as "data field added." Further, it is this error type which is most closely associated with program processing failure, which accounts for 138, or 96.5%, of the 143 "data field added" errors. (This error type is not as serious as some other types, but may result in wasted space. Programming limitations which may have caused this error type were discussed previously in the section on methodology, Section III. C. 1.)

A comparison of Tables 16 and 17 shows that the Santa Cruz records account for 30.1% of all the errors found in the Subject catalog sample. In particular, it may be noted that 62.9% of the errors caused by program processing failure occurred in the Santa Cruz records. (This percentage is somewhat lower than the percentage of program processing errors in the Author/Title catalog attributable to Santa Cruz records; a discussion of where the program failure could have occurred may be found in Section IV. B. 2., which deals with causes of error in the Author/Title catalog.) It may also be noted that Santa Cruz records account for 64, or 49.2%, of the 130 errors in the Subject catalog sample which are due to variant cataloging practice. However, although in the Author/Title catalog the majority of records, in the Subject catalog the Santa Cruz sample contributed only 19 of the 173 errors with "unknown" causes.

3. Seriousness of Errors in the Subject Catalog

As in the Author/Title catalog, errors in the Subject Catalog were categorized as minor, serious and fatal, with the most serious effect being chosen when an error had more than one effect.

Minor errors, as mentioned in Section IV. B. 3., have three effect categories: wasted space, appearance of the entry, and other or unknown. Table 18 displays these errors according to cause and effect. There were 522 minor errors in the Subject catalog sample, or 44.6% of all errors in the Subject sample. (In the Author/Title segment minor errors accounted for 51.5% of all errors.) Unlike the Author/Title catalog, where keying mistakes far outnumbered program processing errors (55.3% and 18.7% respectively) keying mistakes accounted for only a slightly higher percentage (34.7%) of the minor errors than did program processing failure (34.3%) in the Subject sample. Each of these causes was primarily related to a different effect. For instance, 177, or 97.8%, out of 181 keying errors

Type of Error	CAUSE OF ERROR: Keying Error (10)	Variant Cataloging Practice (20)	Program Processing (30)	Processing (40)	Inadequate Design (50)	Unknown (70)	Total
Duplicate data not suppressed (10)	0	0	0	0	0	8	8(2.2)*
Variant form of same data (20)	0	26	0	0	0	0	26(7.2)
Orthographic inaccuracy--transposition (31)	0	1	0	0	0	0	1(0.3)
Orthographic inaccuracy--missing string (32)	17	25	22	4	0	4	72(20.1)
Orthographic inaccuracy--added string (33)	2	1	3	0	0	0	6(1.7)
Orthographic inaccuracy--meaningless string (34)	4	0	0	0	0	1	5(1.9)
Orthographic inaccuracy--missing or added blanks (35)	21	2	1	0	0	0	24(6.7)
Orthographic inaccuracy--incorrect or missing caps (36)	3	0	0	0	0	0	3(0.8)
Orthographic inaccuracy--other misspelling or undetermined (37)	45	3	3	0	0	0	51(14.2)
String improperly used in filing--function term (41)	0	0	0	0	0	0	0(0.0)
String improperly used in filing--dates (42)	1	5	0	0	0	0	6(1.7)
String improperly used in filing--other (46)	0	0	1	0	0	0	1(0.3)
Data field missing (50)	0	0	0	0	0	2	2(0.6)
Data field added (60)	2	1	138	1	1	0	143(39.8)
Inappropriate entry (70)	0	0	0	7	0	0	7(1.9)
Other (80)	0	0	0	0	0	4	4(1.1)
TOTAL	95(26.5)*	64(17.8)	168(46.8)	12(3.3)	1(0.3)	19(5.3)	359(100.0)

*Numbers in parentheses represent percent of total sample.

Table 17

Number and Percent of All Errors in Sample by Type and Cause, Subject Catalog, Santa Cruz Records only

	CAUSE OF MINOR ERRORS						
Effect of Minor Errors	Keying Error(10)	Variant Cataloging Practice(20)	Program Processing(30)	Processing(40)	Inadequate Design(50)	Unknown(70)	Total
Wasted space (40)	3	0	141	95	20	9	268(51.3)*
Appearance only (80)	177	3	5	0	1	10	196(37.6)
Other or unknown (90)	1	1	33	0	21	2	58(11.1)
TOTAL	181(34.7)*	4(0.8)	179(34.3)	95(18.2)	42(8.0)	21(4.0)	522(100.0)

*Numbers in parentheses represent percent of total sample.

Table 18

Number and Percent of All Minor Errors in Sample by Cause and Effect, Subject Catalog

	CAUSE OF MINOR ERRORS						
Effect of Minor Errors	Keying Error(10)	Variant Cataloging Practice(20)	Program Processing(30)	Processing(40)	Inadequate Design(50)	Unknown(70)	Total
Wasted space (40)	2	0	114	8	1	0	125(64.4)*
Appearance only (80)	38	1	4	0	0	0	43(22.2)
Other or unknown (90)	1	0	25	0	0	0	26(13.4)
TOTAL	41(21.2)*	1(0.5)	143(73.7)	8(4.1)	1(0.5)	0(0.0)	194(100.0)

*Numbers in parentheses represent percent of total sample.

Table 19

Number and Percent of All Minor Errors in Sample by Cause and Effect, Subject Catalog, Santa Cruz Records Only

affected the appearance of the entry only; further, these keying errors accounted for 90.3% of all those minor errors affecting appearance. On the other hand, 141, or 78.8%, of program processing problems resulted in wasted space, and these program processing errors accounted for over half (52.6%) of the total number of minor errors causing wasted space.

The Santa Cruz records account for 37.2% of all the minor errors in the Subject catalog, a figure roughly proportionate to the percentage of minor errors in the Author/Title catalog contributed by the Santa Cruz sample (40.8%). From Table 19, however, we can see that program processing problems caused 143, or 73.7% of the 194 minor errors in the Santa Cruz records (keying errors account for only 21.2% of minor errors in these records), thus accounting for 79.9% of all the minor errors caused by programming problems in the Subject catalog.

For the Santa Cruz sample, as for the whole Subject catalog sample, the most common result of program processing errors was wasted space, and conversely, 114, or 91.2%, of the 125 minor errors causing wasted space in the Santa Cruz sample of the Subject catalog were the result of program processing failure.

Serious errors represent 41.8% of the serious errors in the Subject catalog (see Table 1). As may be seen in Table 20, keying errors account for over half (55.3%) of these errors, with variant cataloging practice and program processing the next most frequent causes (20.0% and 16.5%, respectively). It is interesting to note on Table 21 that for the Santa Cruz section of the subject sample, variant cataloging practice accounted for slightly more serious errors than did keying errors (38.8% and 38.1%, respectively). Further, although in the entire Subject catalog sample -- as in the Author/Title catalog -- a large proportion (43.1%) of the serious errors resulted in the effect called "contents uncertain" in the Santa Cruz section of the subject sample, the largest proportion of the errors (41.0%) resulted in non-consolidation of entry in the same column without heading. The "contents uncertain" category accounted for only 20.2% of the serious errors in the Santa Cruz subject sample.

It was pointed out in the discussion of serious errors in Section IV. B. 3. that describing errors in the "content uncertain" category as "serious" was somewhat arbitrary and that it might be therefore useful to review the results of the errors under "contents uncertain" were considered minor instead of serious. Deleting the "contents uncertain" category would reduce the total number of serious errors in the subject sample from 490 to 279. The number and relative proportion of errors due to keying mistakes would be reduced to 124 or 44.4% of the new total. The number of errors due to program processing mistakes would be reduced to 30 (10.8%). On the other hand, although the number of errors due to variant cataloging practice would stay the same, the proportion of these errors relative to the total would be increased to 35.1%. Similarly, in the Santa Cruz portion of this sample, keying errors would be reduced to 29 (33.7%) out of new total of 86 serious errors. The number of errors due to variant cataloging and program processing would change little, but the proportion of variant cataloging practice errors relative to the

Effect of Serious Errors	CAUSE OF SERIOUS ERROR							
	Keying Error(10)	Variant Cataloging Practice(20)	Program Processing(30)	Processing(40)	Inadequate Design(50)	Unknown(70)	Total
Misfiled in same column w/heading (21)	1	0	0	0	0	0	1(0.2)*
Misfiled on same page w/heading (22)	1	0	0	0	0	0	1(0.2)
Content uncertain (30)	147	0	51	0	3	10	211(43.1)
Non-consolidated entry --in same column w/heading (51)	11	6	2	0	4	1	24(4.9)
--on same page w/heading (52)	1	1	0	0	0	0	2(0.4)
--in same column without heading (54)	72	41	23	0	0	10	146(29.8)
--on same page without heading (55)	0	2	0	0	0	0	2(0.4)
Non-consolidated heading --in same column w/heading (61)	5	8	1	0	3	0	17(3.5)
Non-consolidated heading --in same column without heading (64)	1	0	0	0	2	0	3(0.6)
Non-consolidated subject heading: counting subject heading space							
--subject heading non-consolidation in same column (71)	14	12	1	0	0	1	28(5.7)
--subject heading non-consolidation on same page (72)	9	2	1	1	0	2	15(3.1)
--entry heading non-consolidation in same column (74)	0	1	1	0	0	0	2(0.4)
--entry heading non-consolidation on same page (75)	0	1	0	0	0	0	1(0.2)
--whole entry non-consolidation in same column (77)	3	3	1	2	0	0	9(1.9)
--whole entry non-consolidation on same page (78)	1	3	0	0	0	1	5(1.0)
Non-consolidated subject heading: not counting subject heading space							
--whole entry misfile in same column (01)	3	0	-	-	-	-	3(0.6)
--whole entry misfile on same page (02)	0	8	-	-	-	-	8(1.6)
--body only misfile in same column (04)	2	1	-	-	-	-	3(0.6)
--entry heading non-consolidation in same column (08)	0	2	-	-	-	-	2(0.4)
--entry non-consolidation on same page (11)	0	0	-	-	-	-	0
TOTAL	271(55.3)*	98(20.0)	81(16.5)	3(0.6)	12(2.5)	25(5.1)	7(1.4) 490(100.0)

*Numbers in parentheses represent percent of total sample.

Table 20

Number and Percent of All Serious Errors in Sample by Cause and Effect, Subject Catalog

Effect of Serious Errors	CAUSE OF SERIOUS ERROR: Keying Error(10)	Variant Cataloging Practice(20)	Program Processing(30)	Processing(40)	Unknown(70)	Total
Content uncertain (30)	24	0	1	0	3	28(20.2)*
Non-consolidated entry --in same column w/heading (51)	2	1	2	0	0	5(3.6)
--on same page w/heading (52)	1	0	0	0	0	1(0.7)
--in same column without heading (54)	18	24	14	0	1	57(41.0)
--on same page without heading (55)	0	2	0	0	0	2(1.5)
Non-consolidated heading --in same column w/heading (61)	0	0	1	0	0	1(0.7)
Non-consolidated heading --in same column without heading (64)	1	0	0	0	0	1(0.7)
Non-consolidated subject heading: counting subject heading space						
--subject heading non-consolidation in same column (71)	2	5	1	0	1	9(6.5)
--subject heading non-consolidation on same page (72)	2	0	0	1	2	5(3.6)
--entry heading non-consolidation in same column (74)	0	0	1	0	0	1(0.7)
--entry heading non-consolidation on same page (75)	0	1	0	0	0	1(0.7)
--whole entry non-consolidation in same column (77)	1	2	1	2	0	6(4.3)
--whole entry non-consolidation on same page (78)	1	3	0	0	1	5(3.6)
Non-consolidated subject heading: not counting subject heading space						
--whole entry misfile in same column (01)	1	0	-	-	-	1(0.7)
--whole entry misfile on same page(02)	0	6	-	-	-	6(4.3)
--body only misfile in same column(04)	0	1	-	-	-	1(0.7)
--entry heading non-consolidation in same column (08)	0	2	-	-	-	2(1.4)
--entry non-consolidation on same page(11)	0	7	-	-	-	7(5.1)
TOTAL	53(38.1)*	54(38.8)	21(15.1)	3(2.2)	8(5.8)	139(100.0)

*Numbers in parentheses represent percent of total sample

Table 21

Number and Percent of All Serious Errors in Sample by Cause and Effect, Subject Catalog Santa Cruz Records Only

new total for the Santa Cruz subject sample would increase to 62.8%. Program processing would then account for 23.3% of the 86 serious errors. Further, if the "contents uncertain" category were deleted, the effect category "non-consolidated entry in same column without heading" would represent by far the largest proportion of serious errors in the subject sample, accounting for 66.3% of the Santa Cruz section and for 52.3% of the whole Subject catalog sample.

As can be seen in Table 1, the proportion of serious errors to the total number of errors in the Subject catalog is comparable to the proportion of these errors to the total in the Author/Title catalog. There are proportionately fewer minor errors in the Subject catalog (44.6% of the subject sample) than in the Author/Title catalog (51.5%). (As mentioned earlier, far more of these errors are caused by program processing problems in the Subject catalog than are so caused in the Author/Title catalog.) It is interesting to note in Table 22, then, that fatal errors are significantly more numerous in the Subject catalog than in the Author/Title catalog; fatal errors account for 13.6% of the total errors in the subject sample as opposed to accounting for 4.4% of the total errors in the Author/Title catalog. Further, as can be seen in Table 2, the rate of fatal errors per page in the Subject catalog (4.8%) is over twice that of the Author/Title catalog (2.3%). Unfortunately, little can be noted about the factors determining these differences. By far the largest proportion of fatal errors (72.3%) are ascribed to "unknown" causes, a category used whenever there was great uncertainty as to the specific cause of the error. Variant cataloging practice is responsible for the next largest proportion (13.2%). One may note that 114 out of the 115 fatal errors due to "unknown" causes resulted in misfiling off the page. (Such misfiles account for 80.5% of the total fatal errors in the Subject catalog.) In the Author/Title catalog, by contrast, inadequate design was responsible for the plurality of fatal errors (41.1%), with variant cataloging practice (23.4%) and keying mistakes (21.3%) the next most significant causes. For the Subject catalog, inadequate design accounted for only 2.5% and keying erros for only 8.8% of all 159 fatal errors.

As was true of fatal errors in the Author/Title catalog, only a small percentage (9.4%) of the fatal errors in the Subject catalog occurred in in the Santa Cruz records. The number of errors displayed in Table 23 are too small to be of great statistical significance, but one may note that 8 of the 15 fatal errors in the Santa Cruz sample were due to variant cataloging practice, and 4 of the 15 ascribed to unknown causes.

4. Location of Appearance of the Errors in the Subject Catalog

Table 24 displays data correlating the cause of errors in the Subject catalog with the location of the appearance of the errors. Errors in the Subject catalog sample occurred most frequently in the subject heading; 31.7% of the errors in the sample were found in subject headings. Over a third (137, or 36.2%) of the errors in the subject headings were ascribed to "unknown" or uncertain causes; slightly fewer (110, or 29.1%) were due to keying mistakes. The next most frequent location of error appearance was title statement (255, or 21.4% of the total); most of these errors were caused by keying mistakes.

Effect of Fatal Error	CAUSE OF FATAL ERROR: Keying Error (10)	Variant Cataloging Practice (20)	Program Processing (30)	Processing (40)	Inadequate Design (50)	Other (60)	Unknown (70)	Total
Lost entry point (10)	0	0	2	0	0	0	0	2(1.3)*
Misfiled off the page with heading (23)	4	2	0	0	3	0	50	59(37.1)
Misfiled off the page without heading (26)	0	5	0	0	0	0	64	69(43.4)
Non-consolidation of entry off the page with heading (53)	0	1	0	0	0	0	0	1(0.6)
Non-consolidation of entry off the page without heading (56)	0	1	0	0	0	0	0	1(0.6)
Non-consolidated subject heading, counting subject heading space								
--non-consolidation of subject heading off the page (73)	8	5	1	0	1	0	1	16(10.1)
--non-consolidation of entry heading off the page (76)	0	1	0	0	0	0	0	1(0.6)
--non-consolidation of whole entry off the page (79)	2	1	2	0	0	0	0	5(3.2)
Non-consolidated subject heading, not counting subject heading space								
--whole entry misfiled off the page (03)	0	4	0	0	0	0	0	4(2.5)
--non-consolidation of entry off the page (12)	0	1	0	0	0	0	0	1(0.6)
TOTAL	14(8.8)*	21(13.2)	5(3.2)	0	4(2.5)	0	115(72.3)	159(100.0)

*Numbers in parentheses represent percent of total sample.

Table 22

Number and Percent of All Fatal Errors in Sample by Cause and Effect, Subject Catalog

Effect of Fatal Error	CAUSE OF FATAL ERROR: Keying Error (10)	Variant Cataloging Practice (20)	Program Processing (30)	Processing (40)	Inadequate Design (50)	Other (60)	Unknown (70)	Total
Misfiled off the page with heading (23)	0	0	0	0	0	0	4	4(26.7)*
Non-consolidation of entry off the page without heading (56)	0	1	0	0	0	0	0	1(6.7)
Non-consolidation of subject heading counting subject heading space								
--non-consolidation of subject heading off the page (73)	1	1	0	0	0	0	0	2(13.3)
--non-consolidation of whole entry off the page (79)	0	1	2	0	0	0	0	3(20.0)
Non-consolidation of subject heading not counting subject heading space								
--whole entry misfiled off the page (03)	0	4	0	0	0	0	0	4(26.7)
--non-consolidation of entry off the page (12)	0	1	0	0	0	0	0	1(6.7)
TOTAL	1(6.7)*	8(53.3)	2(13.3)	0	0	0	4(26.7)	15(100.0)

*Numbers in parentheses represent percent of total sample.

Table 23

Number and Percent of all Fatal Errors in Sample by Cause and Effect, Subject Catalog, Santa Cruz Records Only

Location of Appearance of Error	CAUSE OF ERROR: Keying Error (10)	Variant Cataloging Practice (20)	Program Processing (30)	Processing (40)	Inadequate Design (50)	Unknown (70)	Total
Subject entry (30)	37	41	16	79	0	7	180(15.1)*
—subject heading (31)	110	70	42	18	1	137	378(31.7)
—entry heading (32)	63	18	5	1	12	12	111(9.3)
—title statement (33)	226	1	16	0	4	8	255(21.4)
—date (34)	3	0	19	0	0	4	26(2.2)
—call no./location code (35)	[illegible]	0	2	0	0	2	23(1.9)
—other (36)	7	0	167	1	41	3	219(18.4)
TOTAL	465(39.0) *	130(10.9)	267(22.4)	99(8.3)	58(4.9)	173(14.5)	1,192(100.0)

*Numbers in parentheses represent percent of total sample.

Table 24

Number and Percent of All Errors in Sample by Location of Error Appearance and Cause, Subject Catalog

Location of Appearance of Error	CAUSE OF ERROR: Keying Error (10)	Variant Cataloging Practice (20)	Program Processing (30)	Processing (40)	Inadequate Design (50)	Other (60)	Unknown (70)	Total
Subject entry (30)	16	24	15	7	0	0	1	63(17.6)*
—subject heading (31)	11	36	10	5	0	0	15	77(21.5)
—entry heading (32)	8	4	4	0	0	0	0	16(4.5)
—title statement (33)	48	0	1	0	0	0	0	49(13.7)
—date (34)	0	0	0	0	0	0	2	2(0.6)
—call no./location code (35)	6	0	1	0	0	0	1	8(2.2)
—other (36)	5	0	137	0	1	0	0	143(39.9)
TOTAL	94(26.2)*	64(17.9)	168(46.9)	12(3.4)	1(0.3)	0	19(5.3)	358(100.0)

*Numbers in parentheses represent percent of total sample.

Table 25

Number and Percent of all Errors in Sample by Location of Error Appearance and Cause, Subject Catalog, Santa Cruz Records Only

In the Santa Cruz sample, however, errors occurred most frequently in entry elements other than those specifically defined. Table 25 shows that almost 40% of the errors in the Santa Cruz sample occurred in such entry elements, and that program processing was responsible for almost all of these errors. Comparing Tables 24 and 25, one can see that Santa Cruz records were responsible for 143 (65.3%) of the 219 errors found in subject entry elements other than those specifically defined and for 137 (82.0%) of the 167 such errors which were ascribed to program processing failure. As discussed in Section IV. B. 4., the Santa Cruz data for errors in added entries in the Author/Title catalog showed similar trends, and, as noted there, these figures reflect a relatively high proportion of added data fields found in Santa Cruz records (see also Table 27).

The data in Tables 26 and 27, which correlate location of appearance of errors with error type, reflects the trends noted in Tables 24 and 25. For example, it may be seen from Table 26 that orthographic inaccuracies account for most of the errors located in the subject heading and title statement (and for 60.2% of the total number of errors in the Subject sample), as might be expected from the information in Table 24 that a substantial number of such errors were caused by keing mistakes.

It may also be noted that outside the orthographic mistakes, the single largest error type in the Subject sample was "data field added" (18.2%), and that most of these errors occurred in subject entry elements other than those specifically defined. Further, a comparison of Tables 26 and 27 shows that the Santa Cruz records were responsible for 143, or and, again, that most of this type of error in the Santa Cruz records were located in entry elements other than the specifically defined categories. It was pointed out in the discussion of error type and cause (Section IV. C. 2.) that "data field added" was the largest single error type in the Santa Cruz subject sample, accounting for almost 40% of the errors in that subset, and is closely associated with program processing failure in that subset. It can be seen from Table 27 that the primary result of this error type in the Subject catalog and from the program processing problems which caused it is wasted space in the addition of unnecessary entry elements.

5. Location of Origin of Errors in the Subject Catalog

Tables 28 and 29 correlate cause of errors in the Subject catalog with the location of the source of the errors. As explained earlier in the discussion of the corresponding tables for the Author/Title catalog (Tables 14 and 15), these tables differ from Tables 24 and 25 in that, whenever the location of the cause of an error differed from the location of the appearance of that error in the Subject entry, it was subtracted from the general subject entry category (used for tallying such errors) and tallied in main entry location where the error originated.

There were 92 errors in the Subject catalog sample for which the source location differed from the location of appearance; 54 (58.7%) of these were from Santa Cruz records. As was true for the Author/Title catalog, these errors were dispersed in various source locations rather than in a significant few. It may be noted, however, that of the 41 errors caused by variant cataloging practice whose source location differed from the location of appearance, 19 (46.3%) originated in the

TYPE OF ERROR

ORTHOGRAPHIC INACCURACY

FIELDS IMPROPERLY USED OR FILLED

Location of Appearance of Error	Duplicate Data not Suppressed (30)	Variant Forms of Same Data (30)	Transposition (31)	Missing String (32)	Added String (33)	Meaningless String (34)	Missing or Added Blanks (35)	Incorrect or Missing Code (36)	Other Misspellings or Undetermined (37)	Function Term(41)	Dates(42)	Associated Title(43)	Non-English Article(45)	Other(46)	Data Field Missing(50)	Data Field Added(60)	Inappropriate Entry(70)	Other(90)	Total
Subject entry (30)	6	7	2	47	8	1	19	0	7	0	0	0	0	0	1	3	78	1	180 (15.1)*
Subject heading (31)	9	22	30	58	86	4	5	0	79	8	9	0	0	1	19	0	15	41	378 (31.7)
Entry heading (32)	1	6	4	22	9	2	10	14	18	5	10	4	1	0	8	1	0	2	111 (9.3)
Title statement (33)	7	0	10	71	26	6	26	11	95	8	0	0	0	0	1	0	0	2	255 (21.4)
Date (34)	0	0	1	1	0	2	0	0	16	8	0	0	0	0	6	0	0	0	26 (2.2)
Call no./Location code (35)	1	0	0	2	2	6	9	0	5	0	0	0	0	0	0	0	0	0	23 (1.9)
Other (36)	0	0	0	0	0	0	1	0	4	0	0	0	0	0	0	213	0	1	219 (18.4)
TOTAL	24 (2.0)*	35 (2.9)	47 (3.9)	301 (16.9)	131 (11.0)	19 (1.6)	70 (5.9)	25 (2.1)	224 (18.8)	5 (0.4)	19 (1.6)	4 (0.3)	3 (0.3)	1 (0.1)	27 (2.3)	217 (18.2)	93 (7.8)	47 (3.9)	1,192 (100.0)

*Number in parentheses represent percent of total sample

TABLE 76

Numbers and Percent of All Errors in Sample by Type and Location of Error Appearance, Subject Catalog

76

TYPE OF ERROR

Location of Appearance of Error	Duplicate Data not Suppressed (10)	Variant Forms of Same Data(20)	ORTHOGRAPHIC INACCURACY (31)-(37) Transposition(31)	Missing String (32)	Added String (33)	Meaningless String(34)	Missing or Added Blanks(35)	Incorrect or Missing Caps(36)	Other Misspellings or Undetermined (37)	STRING IMPROPERLY USED IN FILING Dates (42)	Other (46)	Data Field Missing (50)	Data Field Added(60)	Inappropriate Entry (70)	Other (80)	Total
Subject entry (30)	0	5	0	33	1	0	12	0	3	0	0	0	3	6	0	63(.716)*
Subject heading (31)	8	21	1	30	0	1	1	0	6	3	1	0	0	1	4	77(21.5)
Entry heading (32)	0	0	0	2	3	0	2	1	5	3	0	0	0	0	0	16(4.5)
Title statement (33)	0	0	0	7	2	0	9	2	29	0	0	0	0	0	0	49(13.7)
Date (34)	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	2(0.6)
Call no./Location code (35)	0	0	0	0	0	4	0	0	4	0	0	0	0	0	0	8(2.2)
Other (36)	0	0	0	0	0	0	0	0	3	0	0	0	140	0	0	143(39.9)
TOTAL	8(2.2)*	26(7.3)	1(0.3)	72(20.1)	6(1.7)	5(1.4)	24(6.7)	3(0.8)	50(14.0)	6(1.7)	1(0.3)	2(0.6)	143(39.9)	7(1.9)	4(1.1)	358(100.0)

*Numbers in parentheses represent percent of total sample

Table 27

Number and Percent of All Errors in Sample by Type and Location of Error Appearance, Subject Catalog, Santa Cruz Records Only

Location of Source of Error	CAUSE OF ERROR: Keying Error (10)	Variant Cataloging Practice (20)	Program Processing (30)	Processing (40)	Inadequate Design (50)	Other (60)	Unknown (70)	Total
Main entry--entry heading (11)	0	1	0	0	0	0	1	2(0.2)*
--title statement, short title (41)	6	2	0	0	0	0	0	8(0.7)
--title statement, subtitle (42)	12	1	1	0	0	0	0	14(1.2)
--title statement, author statement (43)	2	19	1	0	0	0	1	23(1.9)
--title statement, other (44)	4	3	0	0	0	0	0	7(0.6)
--edition statement (13)	1	5	2	0	0	0	0	8(0.7)
--place (14)	0	2	1	0	0	0	0	3(0.2)
--publisher (15)	6	1	1	0	0	0	0	8(0.7)
--date (16)	0	3	7	0	0	0	0	10(0.8)
--collation (17)	4	3	0	0	0	0	0	7(0.6)
--notes (18)	1	1	0	0	0	0	0	2(0.2)
Subject entry (30)	1	0	3	79	0	0	5	88(7.4)
--Subject heading (31)	110	70	42	18	1	0	137	378(31.7)
--entry heading (32)	63	18	5	1	12	0	12	111(9.3)
--title statement (33)	226	1	16	0	4	0	8	255(21.4)
--date (34)	3	0	19	0	0	0	4	26(2.2)
--call no./location code (35)	19	0	2	0	0	0	2	23(1.9)
--other (36)	7	0	167	1	41	0	3	219(18.3)
TOTAL	465(39.0)*	130(10.9)	267(22.4)	99(8.3)	58(4.9)	0	173(14.5)	1,192(100.0)

*Numbers in parentheses represent percent of total sample.

Table 28

Number and Percent of All Errors in Sample by Location of Error Source and Cause, Subject Catalog

Location of Source of Error	CAUSE OF ERROR: Keying Error (10)	Variant Cataloging Practice (20)	Program Processing (30)	Processing (40)	Inadequate Design (50)	Other (60)	Unknown (70)	Total
Main entry								
--title statement, short title (41)	4	1	0	0	0	0	0	5(1.4)*
--title statement, subtitle (42)	6	1	1	0	0	0	0	8(2.2)
--title statement, author statement (43)	0	11	1	0	0	0	1	13(3.6)
--title statement, other (44)	1	2	0	0	0	0	0	3(0.8)
--edition statement (13)	0	3	2	0	0	0	0	5(1.4)
--place (14)	0	2	1	0	0	0	0	3(0.8)
--publisher (15)	5	1	1	0	0	0	0	7(2.0)
--date (16)	0	2	7	0	0	0	0	9(2.5)
--collation (17)	0	1	0	0	0	0	0	1(0.3)
Subject of entry (30)	0	0	2	7	0	0	0	9(2.5)
--subject heading (3	11	36	10	5	0	0	15	77(21.5)
--entry heading (32	8	4	4	0	0	0	0	16(4.5)
--title statement (33)	48	0	1	0	0	0	0	49(13.7)
--date (34)	0	0	0	0	0	0	2	2(0.6)
--call no./location code (35)	6	0	1	0	0	0	1	8(2.2)
--other (36)	5	0	137	0	1	0	0	143(40.0)
TOTAL	94(26.3)*	64(17.9)	168(46.9)	12(3.3)	1(0.3)	0	19(5.3)	358(100.0)

*Numbers in parentheses represent percent of total sample.

Table 29

Number and Percent of All Errors in Sample by Location of Error Source and Cause, Subject Catalog, Santa Cruz Records Only

author part of title statements in main entries. In the Santa Cruz records, 11 (45.8%) of 24 such errors were traced to the author part of the title statement. A similar trend was noted earlier for added entries in the Author/Title catalog.

6. Other Correlations of Errors in the Subject Catalog

As for errors in the Author/Title catalog sample, aspects of errors in the Subject catalog have now been correlated and discussed in a number of ways. Error cause has been correlated with three other aspects: type, location (both location of origin and of appearance), and effect. Error type has been correlated with cause and location of appearance. Error effect has been correlated with cause in tables describing minor, serious, and fatal errors, but not with either location or type. However, the source data is available to permit this to be done at a later date if desired.

7. Effect of Non-Consolidation of Entries in the Subject Catalog

As mentioned in the analysis of the Author/Title catalog, time limitations prevented study of the wasted space in UCUCS. The rough estimate of 3 column-inches per page given in that section was based on a sample from the Author/Title catalog. More detailed breakdown and analysis of the non-consolidation and wasted space categories of the Subject catalog would be necessary before a reliable estimate could be made of the space wasted in the Subject catalog. Such analysis cannot be done here, again due to time limitations, but it is to be hoped that some future studies may deal with this problem.

V. REFERENCES

1. University of California Union Catalog of Monograph Cataloged by the Nine Campuses from 1963 through 1967. Berkeley, Institute of Library Research, University of California, 1972.

2. Nancy Slanger, Input Conversion Keying Manual. Unpublished manuscript for use by a Library School class. Spring, 1972.

3. Liz Gibson, "BIBCON-A General Purpose Software System for MARC-Based Book Catalog Production," Journal of Library Automation. Volume 6, Number 4, December 1973, pp. 237-256.

4. Brett Butler, "Automatic Format Recognition of MARC Bibliographic Elements: A Review and Projection," Journal of Library Automation. Volume 7, Number 1, March 1974, pp. 27-42.

5. U. S. Library of Congress, Subject Cataloging Division, Subject Headings Used in the Dictionary Catalogs of the Library of the Congress from 1897 to June 1964, seventh edition. Washington, 1966.

6. Kelley L. Cartwright, "Mechanization of Library Filing Rules." In: Advances in Librarianship, 1970, v. 1. New York and London: Academic Press, 1970, pp. 59-94.

7. Joe E. Hewitt, "Sample Audit of Cards from a University Library Catalog," College and Research Libraries, Volume 33, Number 1, January 1974, pp. 24-27.

8. Nora Cambier, Barton Clark, Robert Daugherty, and Mike Gabriel, "Books in Print 1969: An Analysis of Errors." American Libraries, Volume 1, Number 9, October 1970, pp. 901-902.

9. J. L. Dolby and H. L. Resnikoff, "Efficient Automatic Error-Detection in Processing Bibliographic Records." In: J. L. Dolby, V. Forsyth, and H. L. Resnikoff, An Evaluation of the Utility and Cost of Computerized Library Catalogs. [Washington, D. C.]: U.S. Department of Health, Education and Welfare, 1968, pp. 83-99.

10. Rand Corporation, A Million Random Digits with 100,000 Normal Deviates. Glencoe, Illinois, Free Press, 1955.

11. Douglas Ferguson, Design for a Study of Selected Error and Entry Duplication in the University of California Union Catalog Supplement. Unpublished paper, May 1974.